ESCAP/WMO TYPHOON COMMITTEE ROVING SEMINAR 2023

June, 28 -30th 2023

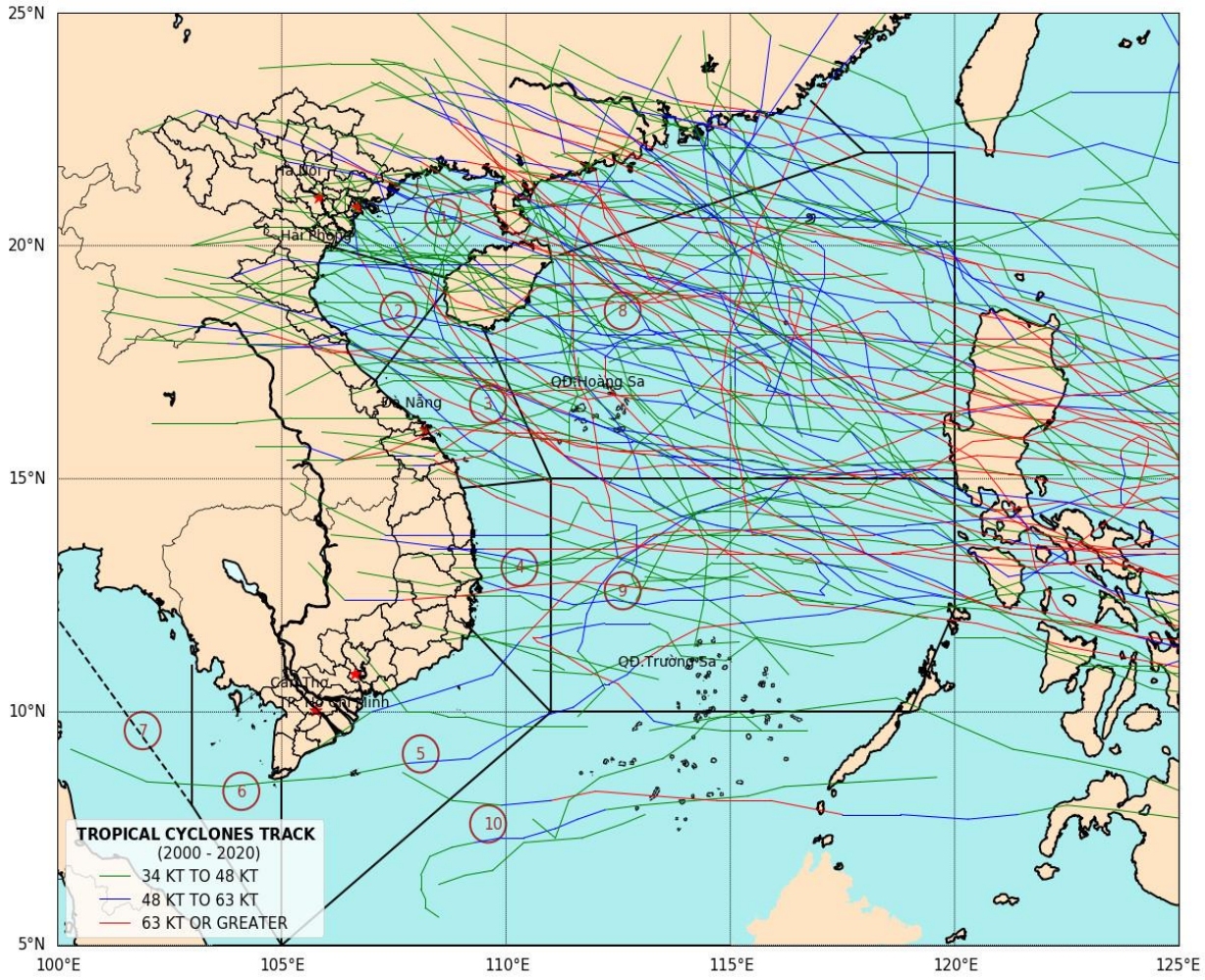
TC-INDUCED RAINFALL ANALYSIS AND FORECASTING IN VIETNAM BY USING SATELLITE DATA AND SELECTIVE ENSEMBLE TECHNIQUE

Author: Assoc. Prof. Pham Thi Thanh Nga
Vietnam Institute of Meteorology, Hydrology, and Climate Change (IMHEN)
Tel: +84 945803088; Email: phamttnga@imh.ac.vn; pttnga.monre@gmail.com

CONTENTS

INTRODUCTION

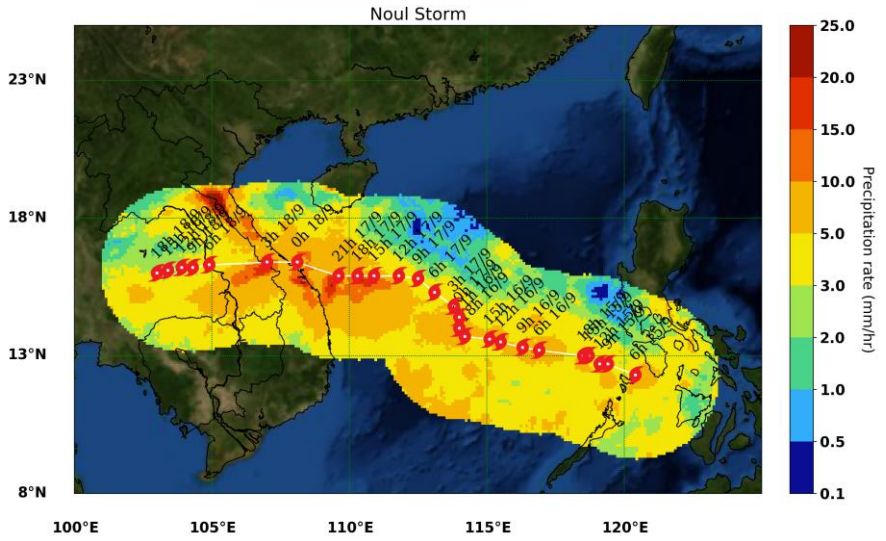
- Vietnam is located in the tropical monsoon region, one of the six TC regions.
- Annually, on average there are about 5-6 storms making landfalls in Vietnam (Trinh et. al., 2014).
- Natural disasters in 2022 have made 175 people dead or missing, causing economic losses of nearly VND 19,500 billion (by the National Steering Committee for Natural Disaster Prevention and Control)
- Impact from tropical cyclones (TCs) and potential induced floods is one of the major natural hazards to the coastal regions of Vietnam.
- Forecasting TC is always a big challenge: The biggest challenge is to accurately determine TC-induced rainfall, which accounts for a large contribution of the annual total rain in the central region of Vietnam

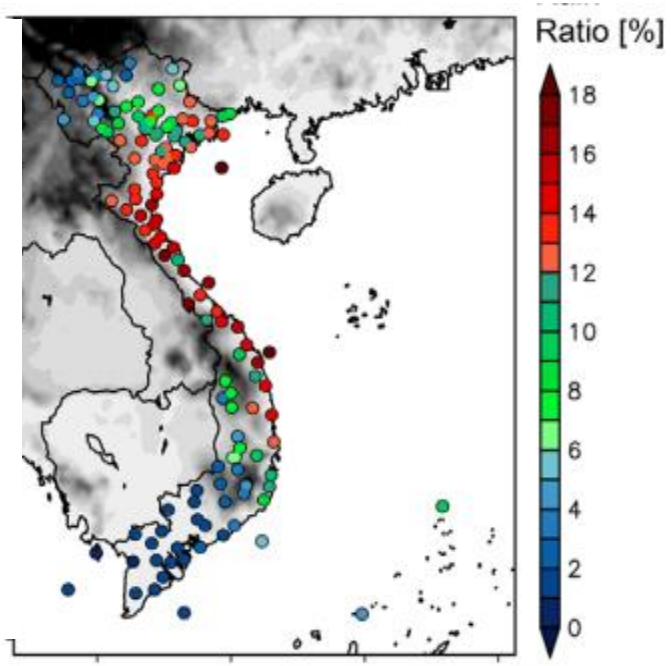


DATA

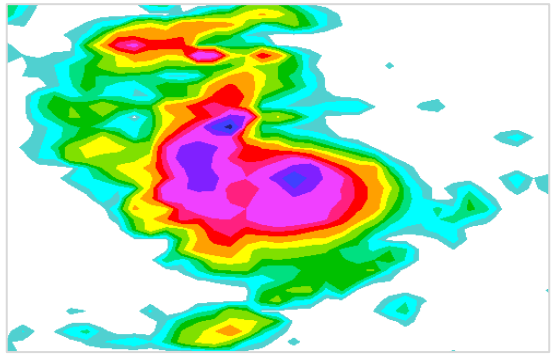
Using data of TCs during 2000-2022 period:

- TC best-track from Vietnam Metoffice
- Rainfall satellite data: GSMAP (0.1x0.1), TRMM (0.25x0.25)
- Rainfall model data: ECMWF 51 Member
- Daily Rainfall Data from 169 meteorological stations

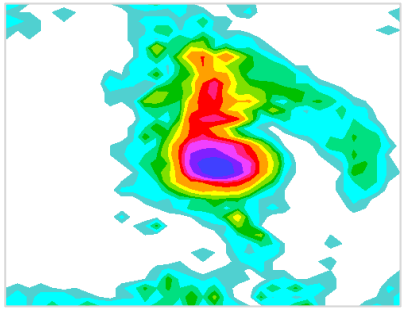




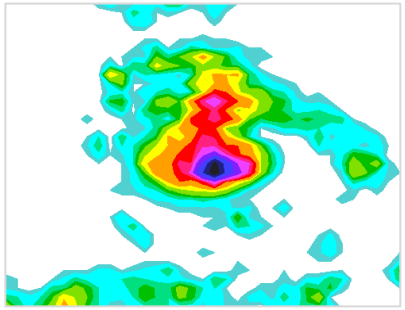
GSMAP 202209250000_06



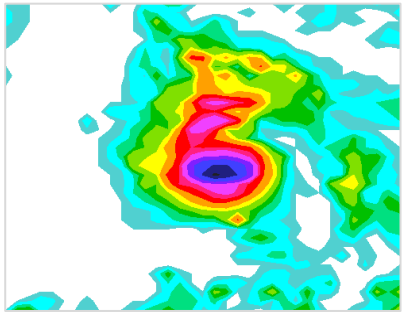
Member 1



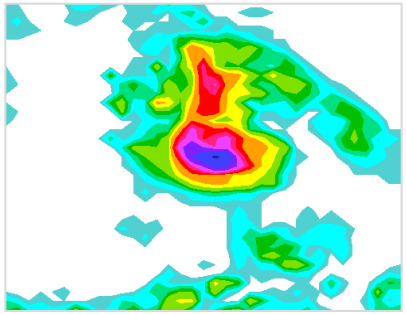
Member 2



Member 3



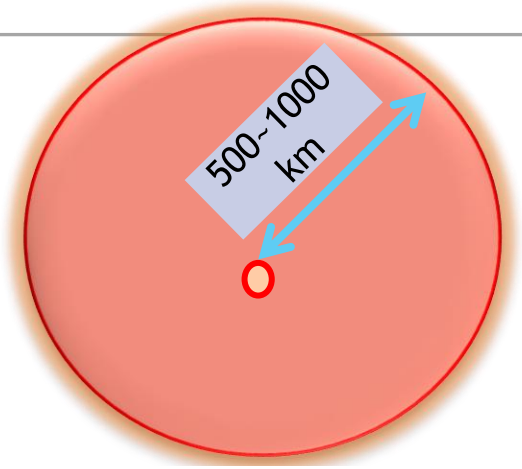
Member 4



ECMWF Member 51

3.1 Evaluation of TRMM/GSMaP data for TC-induced rainfall over Vietnam region

METHOD Detect the Tropical Cyclone Induced Rainfall

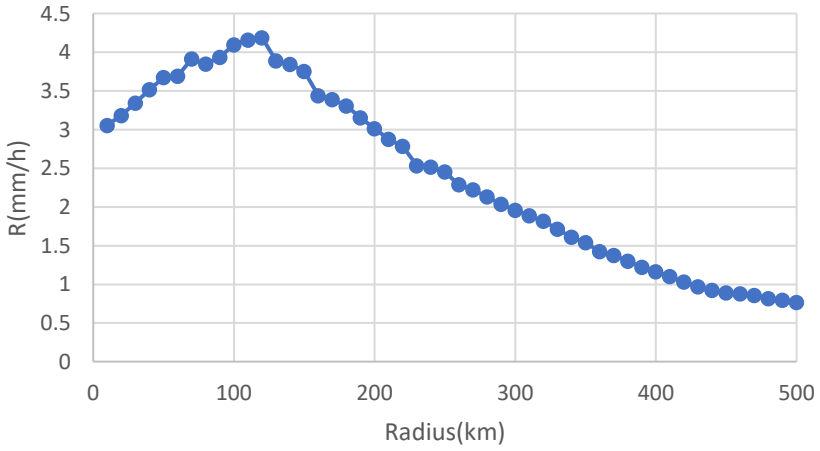


- **TC rainfall could be defined by using distance from the center of the TC.** (Rao and MacArthur 1994; Rodgers et al. 2000; Nguyen-Thi et al. 2012a).
- **The selected radius are common from commonly from 500~1000 km.** (1000km Kubuta (2009); 600km Gleason (2006) Englehart (2001), Hoang Anh (2011)).

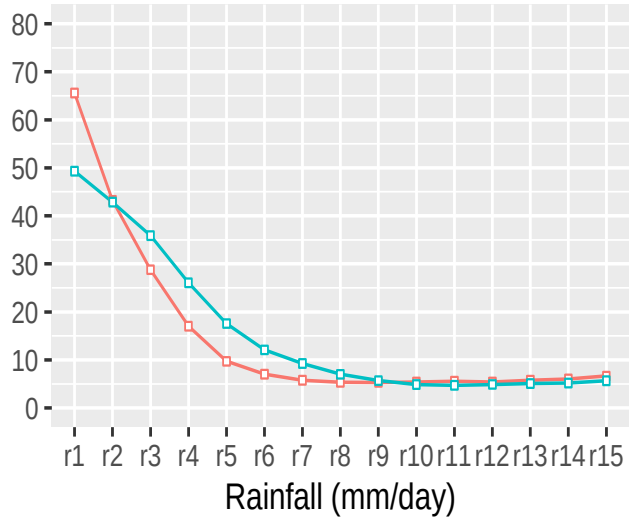
➢ **WE USE 500 KM Radius for whole this study**

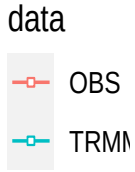
TC-induced rainfall is rains occurring within 500 km of station

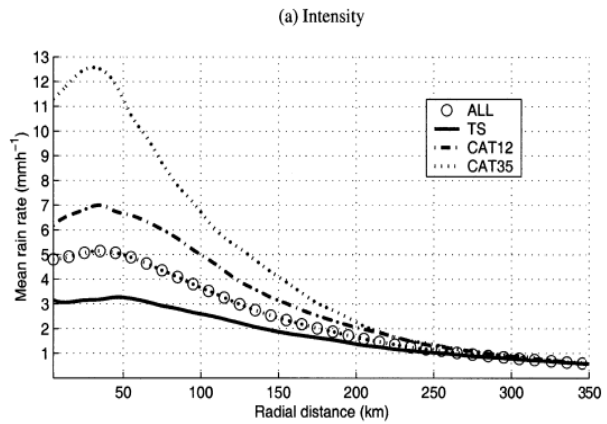
By GSMaP Rain rate(mm/h)

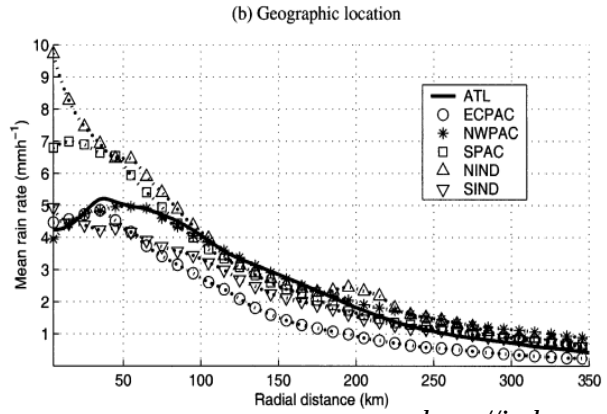


Mean radial distribution of TC rainfall









Evaluate of TRMM/GSMaP Precipitation Estimates during the TC events

The evaluation is applied to 2 cases of classification:

1) Divided by TC intensity (3 categories: TS, STS, TYP)

2) Divided by effected regions (5 regions as in the figure)

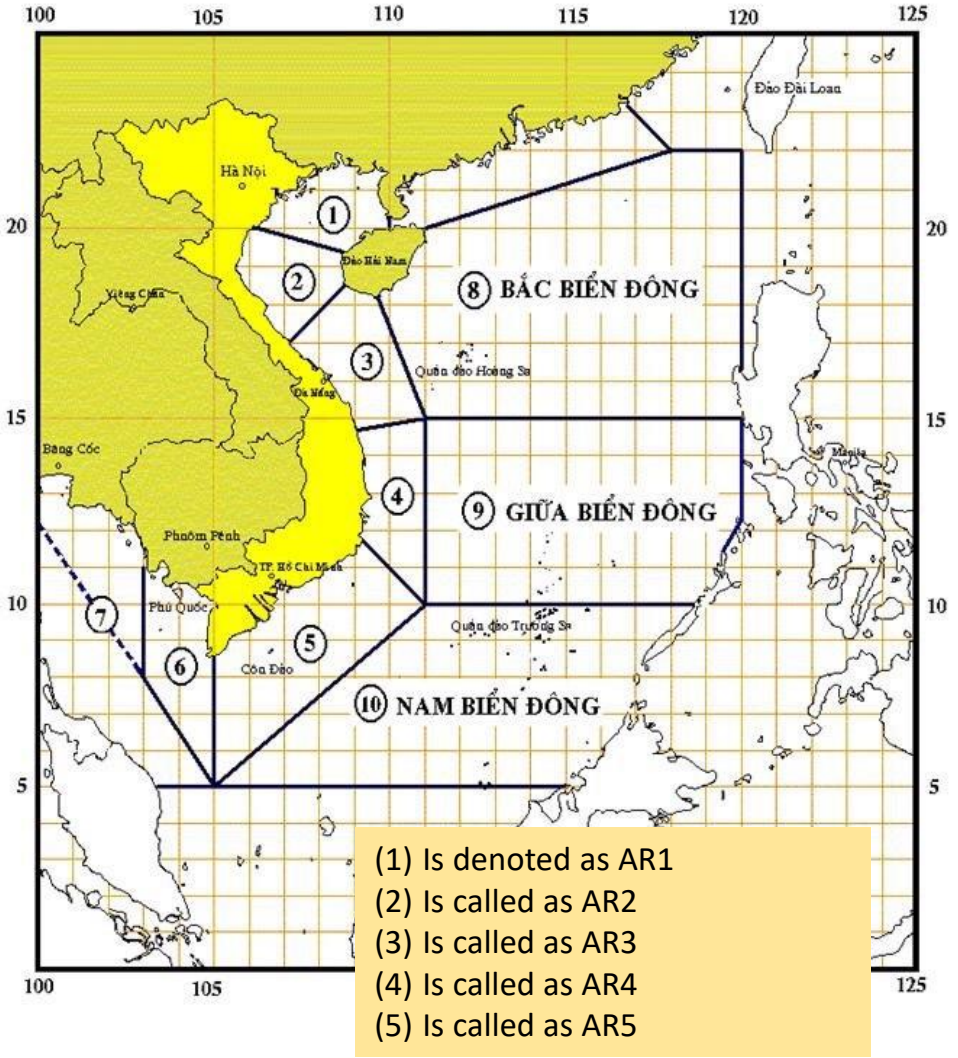
In terms of distance from TC core and position respect to TC track (left and right)

- Continuous Statistics: bias, the ME and MAE
- Categorical statistics: POD, FBI, TS, ETS using a contingency tabl

Rainfall categories (mm day^{-1})	Rainfall threshold (mm day^{-1})
5–15	5
15–30	15
30–45	30
45–75	45
75–100	75
>100	100

3 TC categories:

- TS: (62-88km/h);
- STS (89-117km/h);
- TYP (>=118km/h)



(1) Is denoted as AR1
(2) Is called as AR2
(3) Is called as AR3
(4) Is called as AR4
(5) Is called as AR5

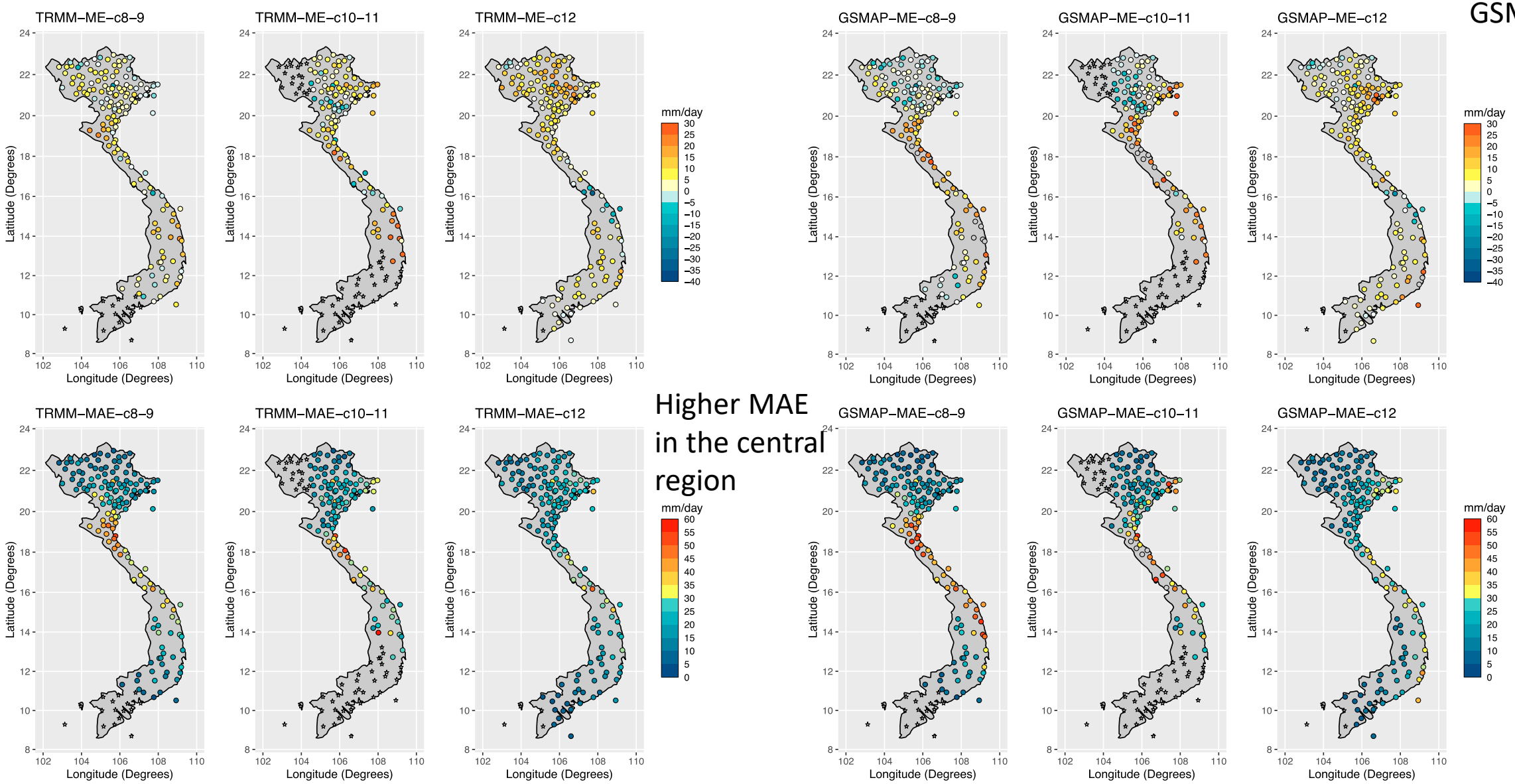
division of **5 landing regions**

RESULTS

Spatial distribution of ME, MAE divided by TC intentisty

TRMM

GSMaP

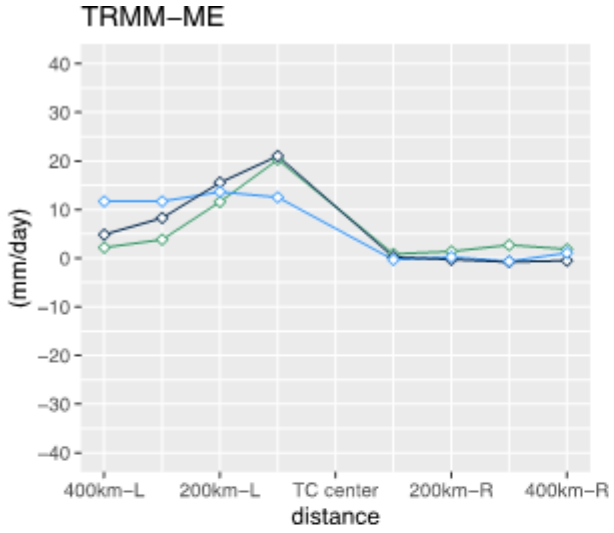


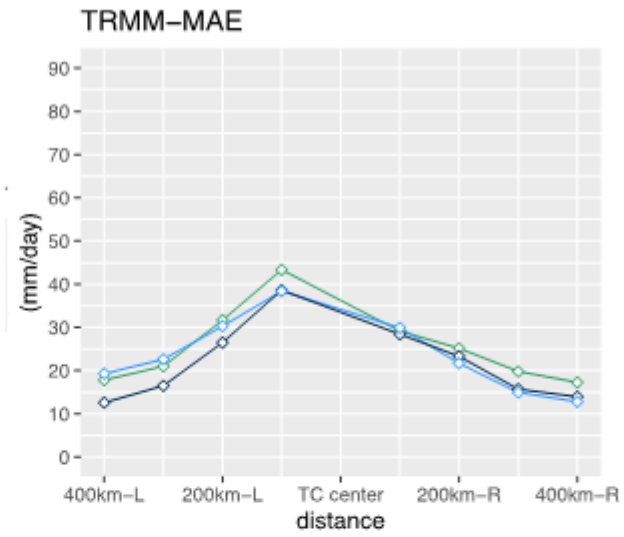
Higher MAE in the central region

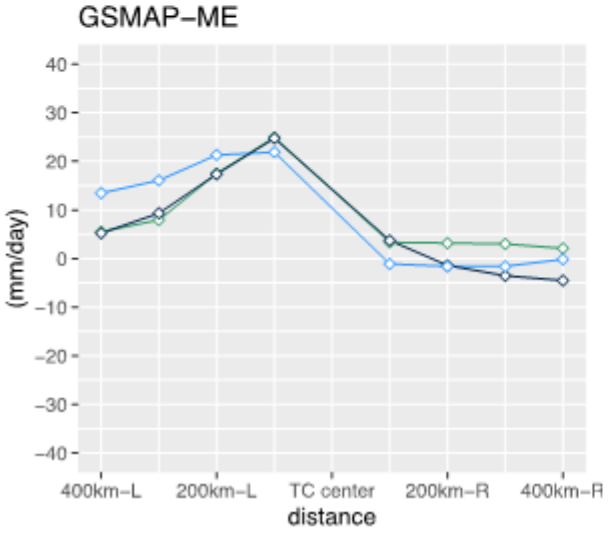
(Stars denoted as the NA value)

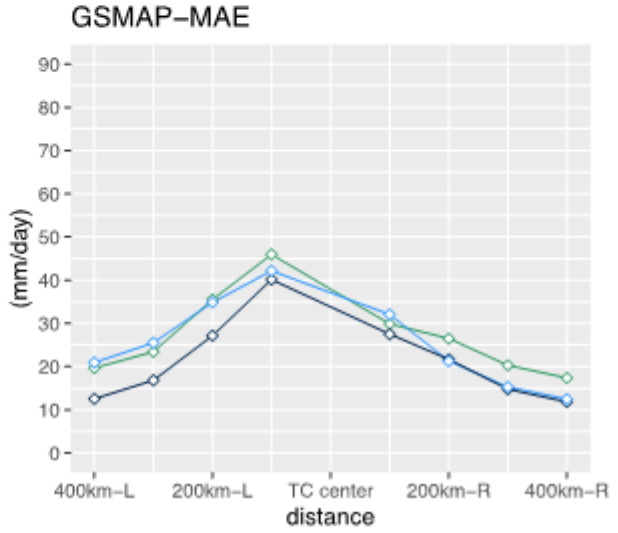
RESULTS

(Radial distribution of ME, MAE in 3 TC types)



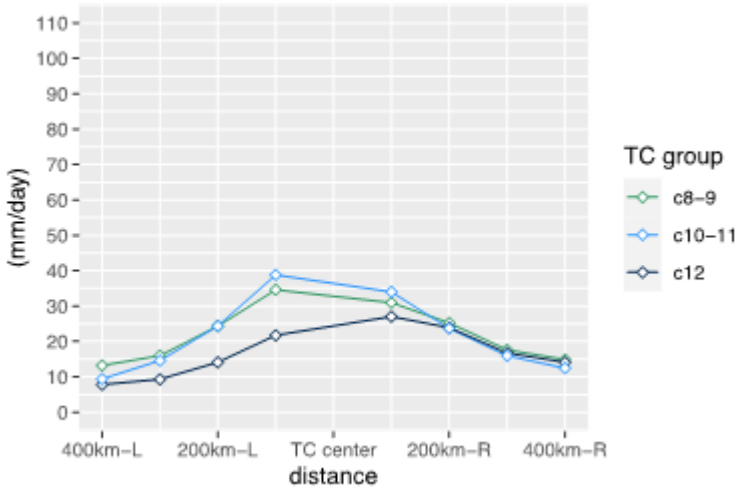






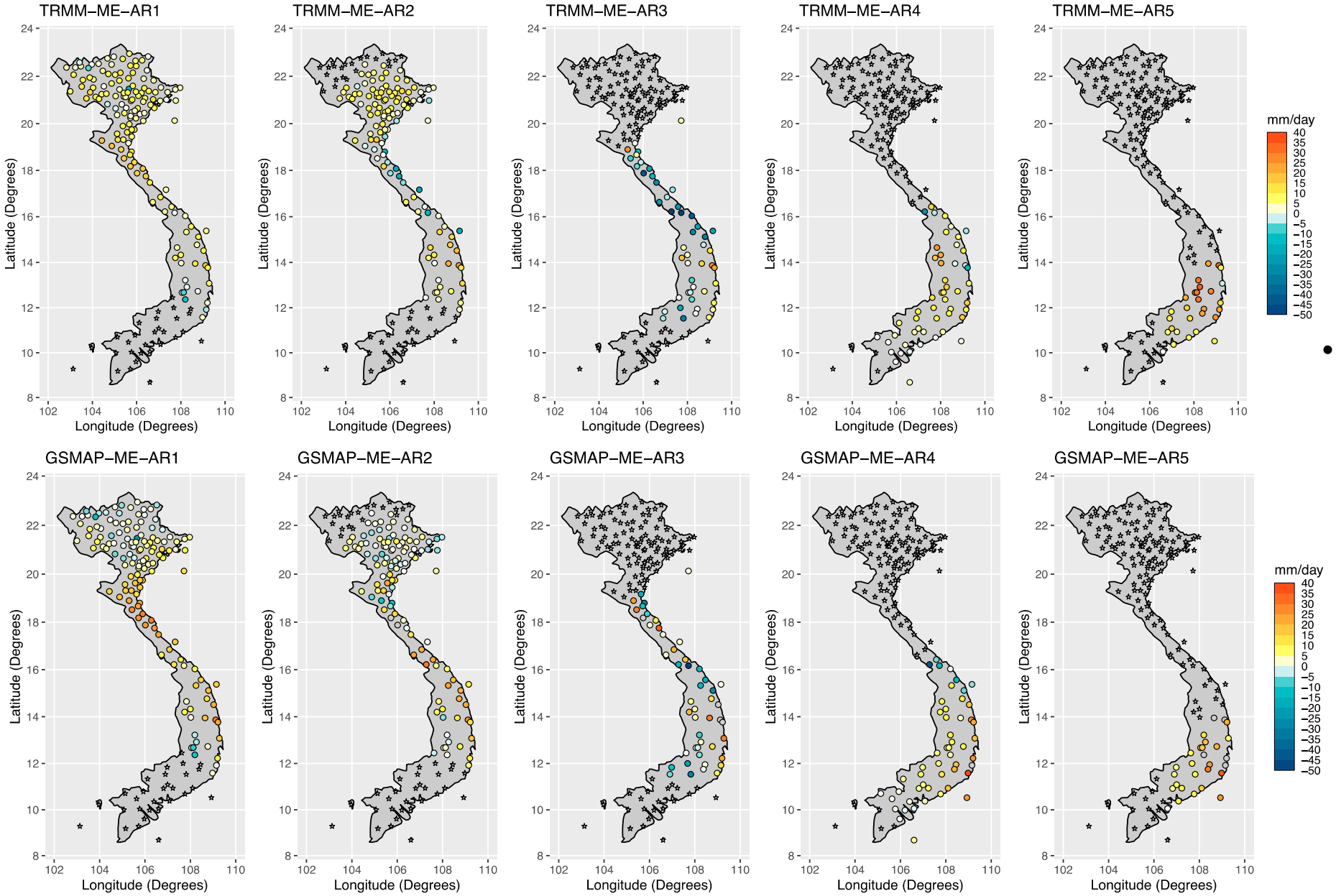
- Overall, estimasted rainfall are over estimasted in the regions within 500km from TC core.
- The highest values of ME, MAE are within 200km from the TC core in the left side

Rainfall distribution by observations



RESULTS

Evaluation separated for 5 TC-landed regions

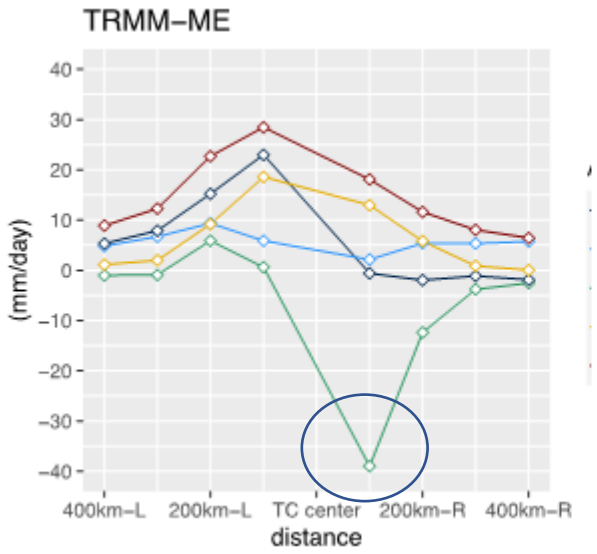


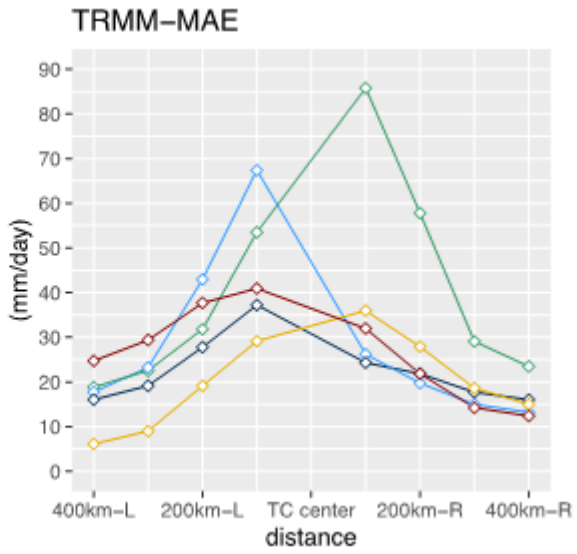
- TRMM and GSMAP are over estimasted for almost station in 5 TC landed regions

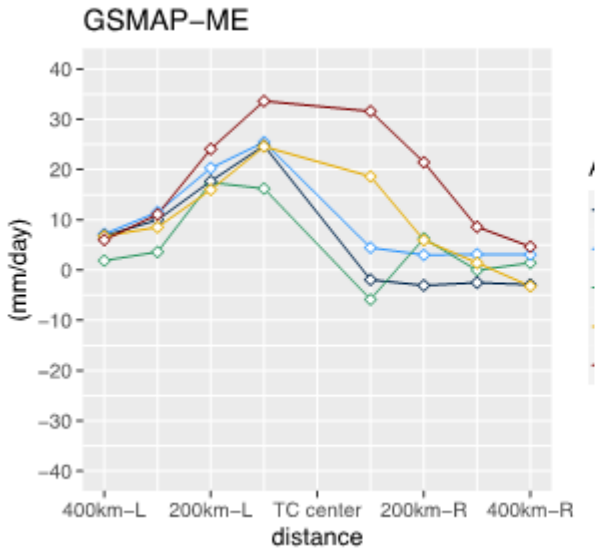
(Stars denoted as the NA value)

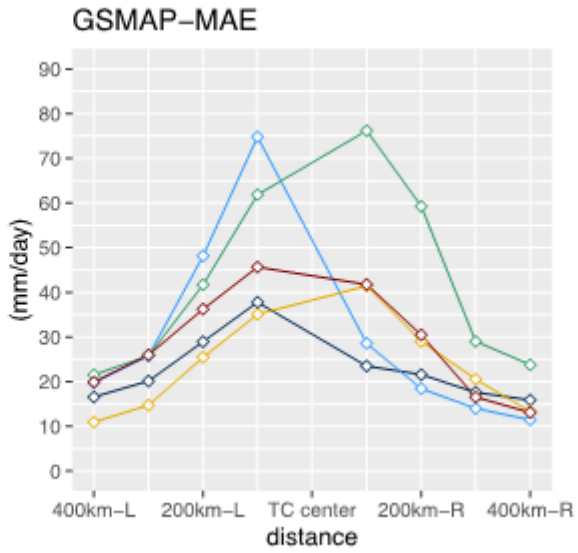
RESULTS

(Radial distribution of ME. MAE in 5 affected regions)

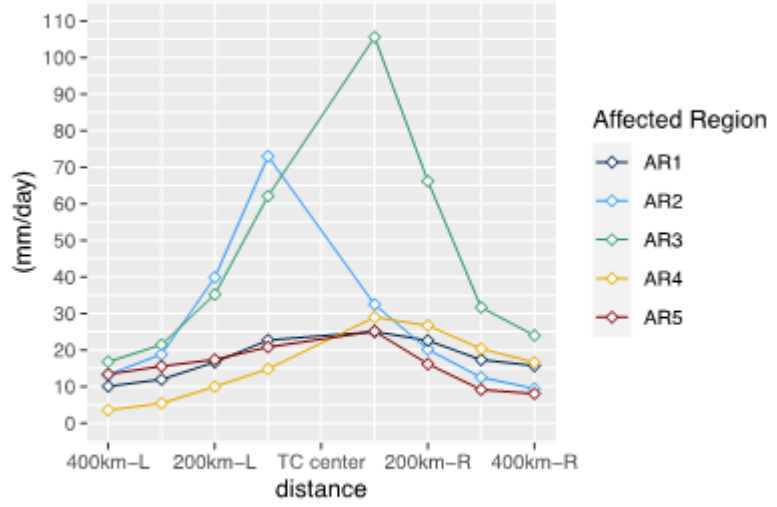




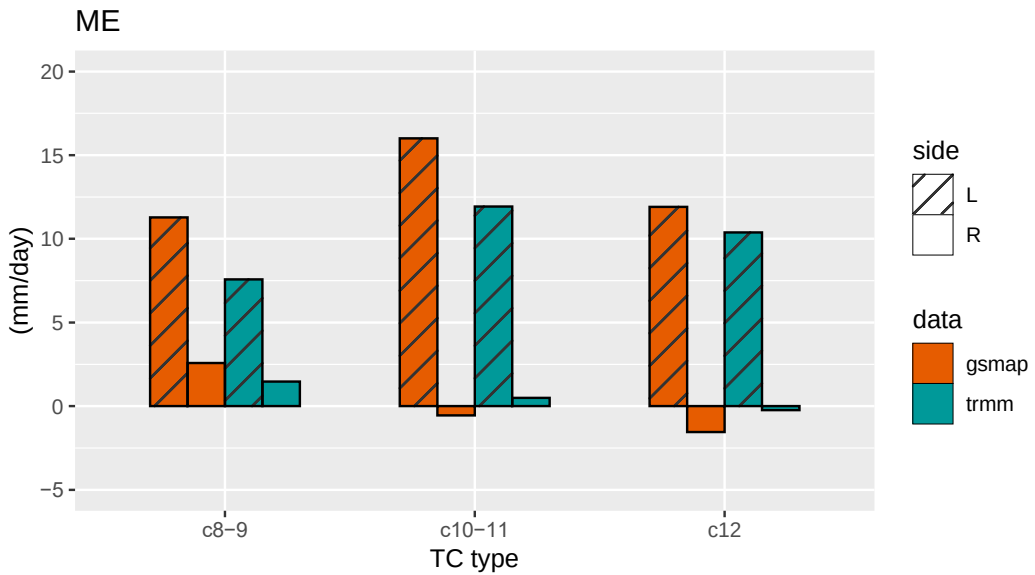


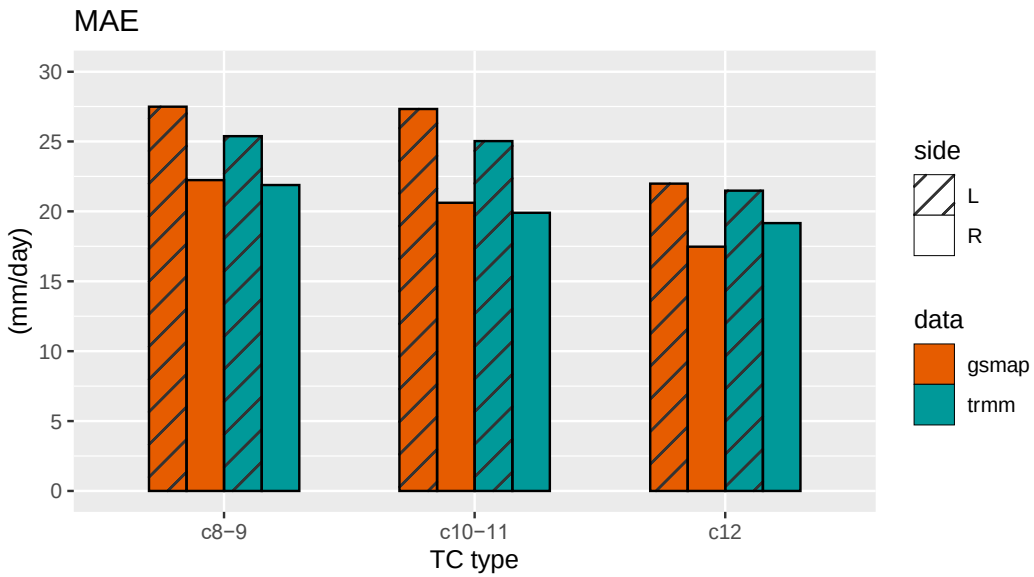


- For both data, the high bias are concentrated around the TC core, which are significantly decreased for the far distance.
- TRMM and GSMAP are over estimasted in the regions within 500km from TC core, except for the radius of 200km in the right side of TC event effecting to the AR3.
- In both data, for TC cases landing in the AR1, AR2, AR5, MAEs tend to higher in the left
- For TC cases landing in the AR3, AR4, the side with high MAE coincides with the high TCR side.

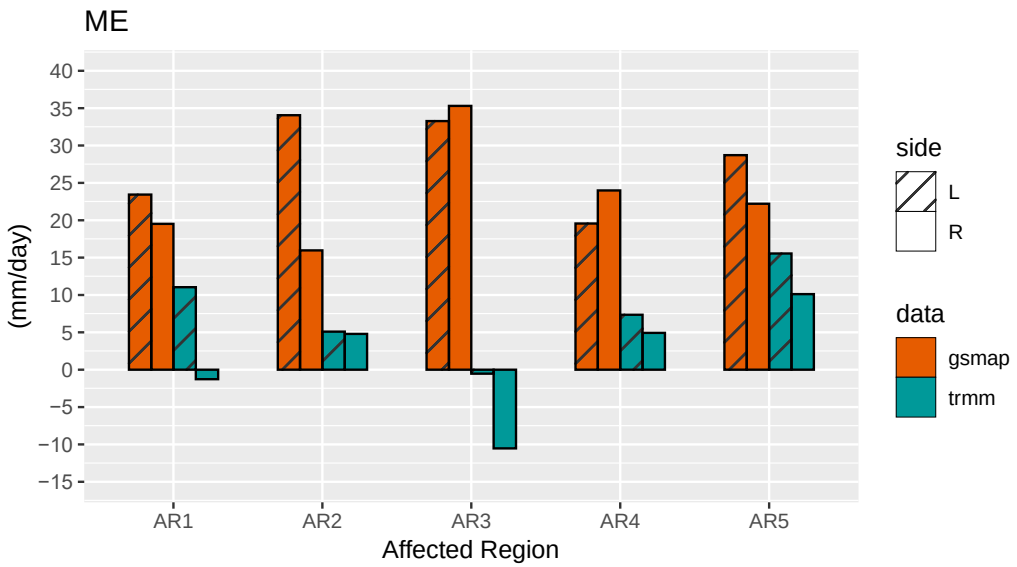


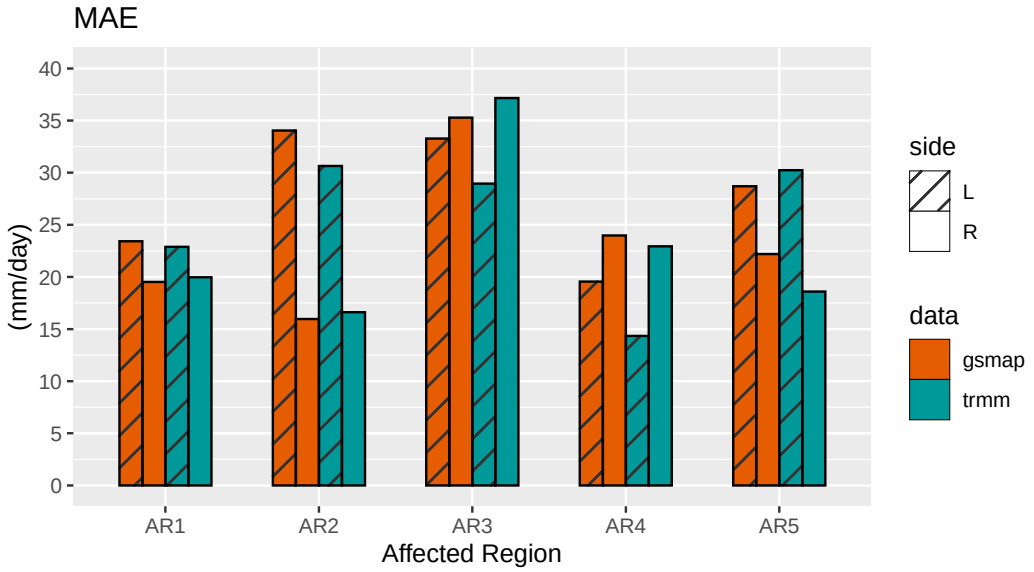
(Compare the value of ME, MAE between the left side and right side of TC in 3 TC categories)





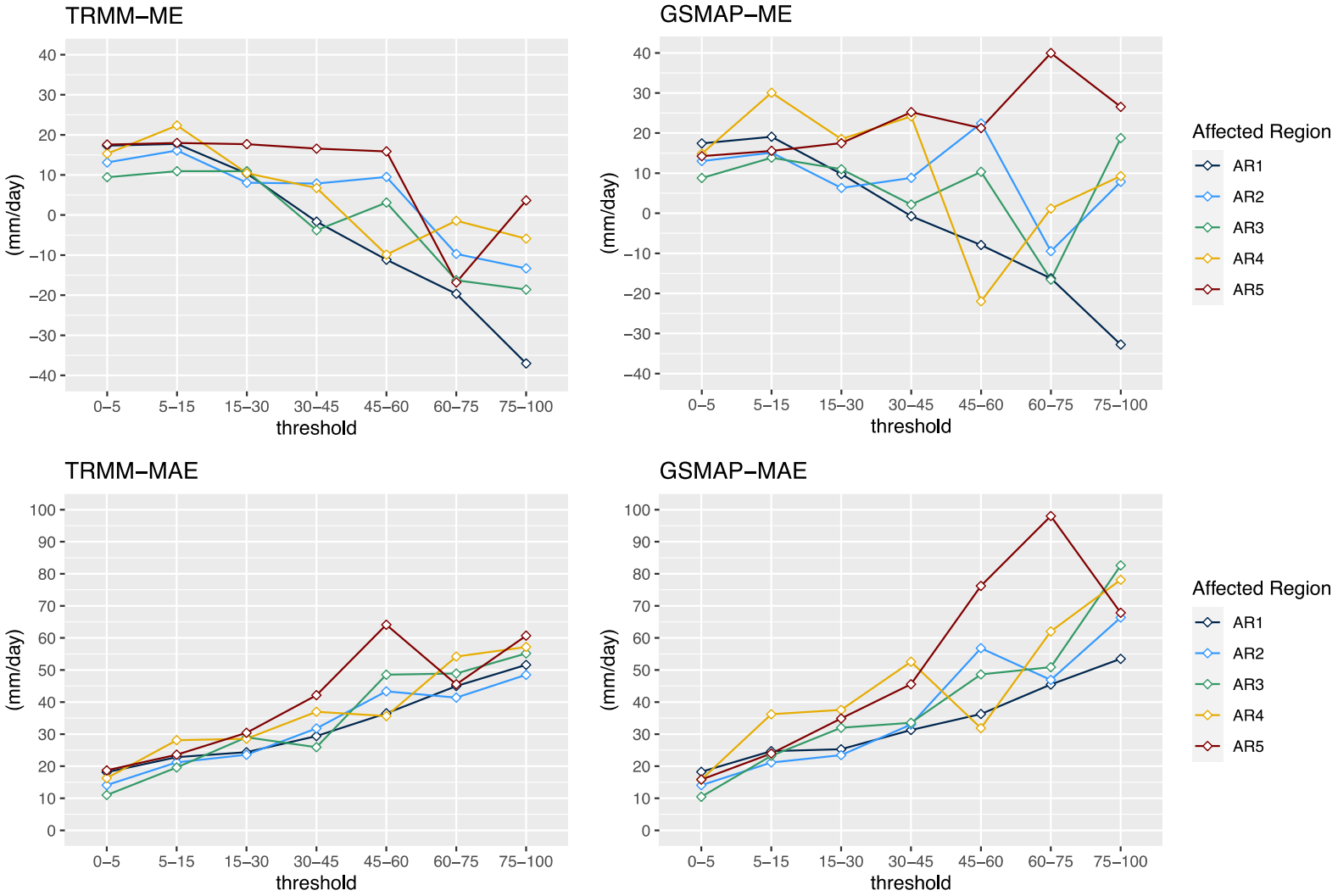
(Compare the value of ME, MAE between the left side and right side of TC in 5 affected regions)





RESULTS

(The ME, MAE at different thresholds on 5 TC regions

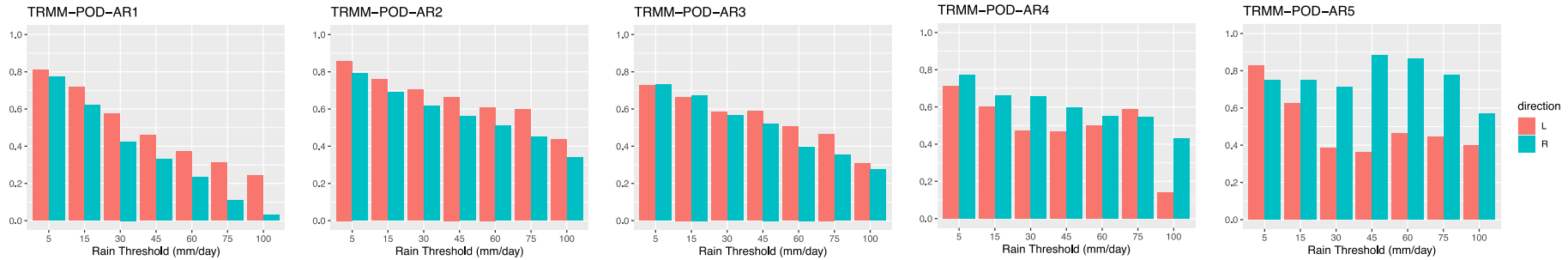


- Overall, both of TRMM and GSMAP tend to overestimate at small rainfall threshold and underestimated at high rainfall threshold for almost affected regions. Except for GSMAP in cases of AR5, ME values are positive for all thresholds.
- The MAE values tend to increase at the high thresholds for both data.

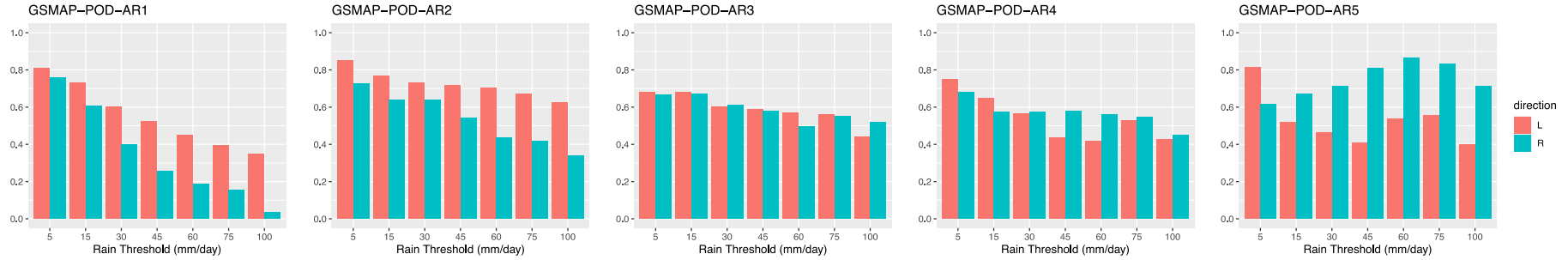
RESULTS

POD

TRMM

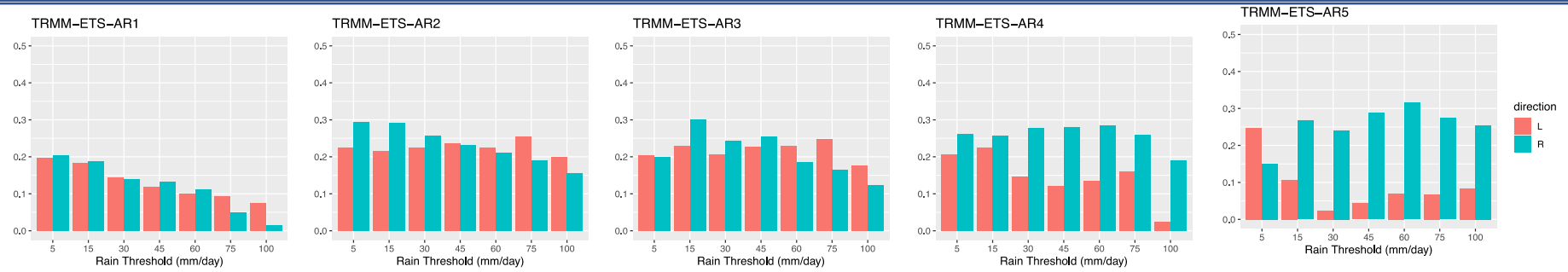


GSMAP

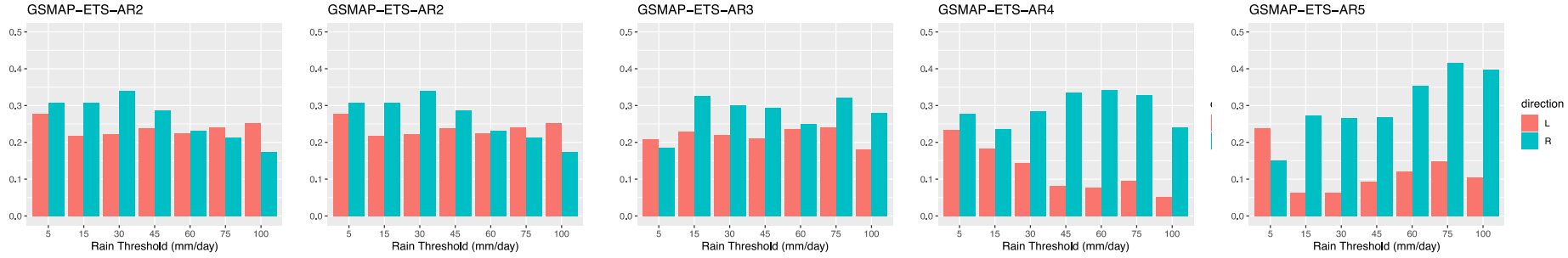


ETS

TRMM



GSMAP



REMARKS 1

- TRMM and GSMAP captures well the radial distribution of TCR, concentrated within 500km radius from TC center.
- Overall, estimasted rainfall are over estimasted in the regions within 500km from TC core. Higher values of ME, MAE are concentrated in the high TCR regions within 200km radius in the left side.
- The MAE values tend to increase at the high thresholds for both data.
- For both of TRMM and GSMAP data, there are the significant difference about the distribution and magnitude of bias depending on landing regions. TRMM and GSMAP are over estimasted in the regions within 500km from TC core.
- In both data, for TC landing in the AR1, AR2, AR5, MAEs is higher in the left, where the rain is lower. While for TC landing in the AR3, AR4, high MAE is in higher TCR side.

3.2 Characteristics of rainfall distribution induced by TCs making landfall over Vietnam using GSMAP

METHOD Analytical method for characteristics of rainfall distribution

Using Fourier decomposition:

The axisymmetric rainfall component is the azimuthally mean rain rate (WN-0) as a function of the radial distance in 10-km-wide annuli from the TC center.

The asymmetric rainfall components is analyzed by binning rainfall rate data in 10-km-wide annuli from each TC center to 500-km radius.

First, the WN-1 Fourier coefficients were computed by using all rain rate estimates as in Boyd:

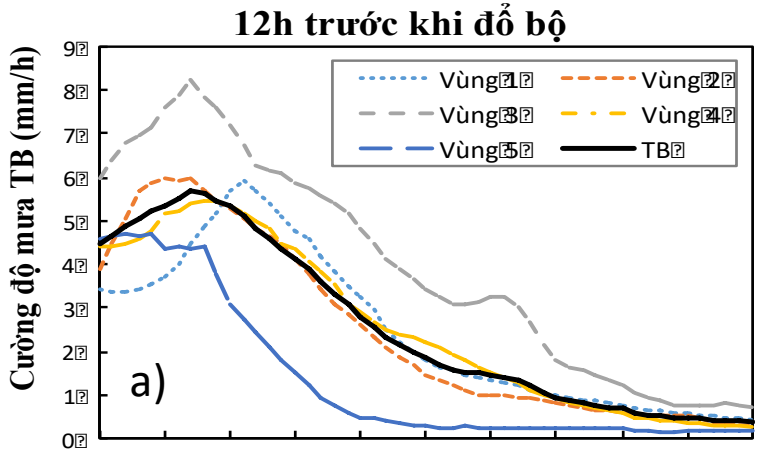
$$a_1 = \sum_i [R_i \cos(\theta_i)], b_1 = \sum_i [R_i \sin(\theta_i)],$$

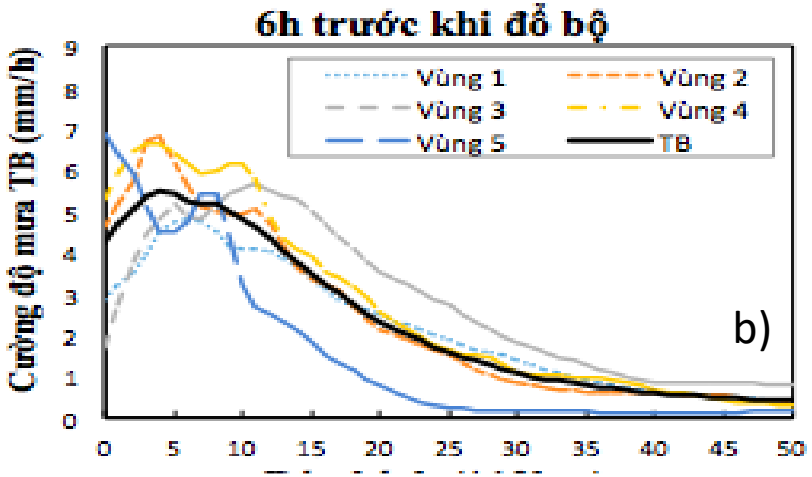
Where R_i is each of the individual rain rate estimate and θ_i is the phase angle of the estimate relative to either the storm motion or the VWS vector

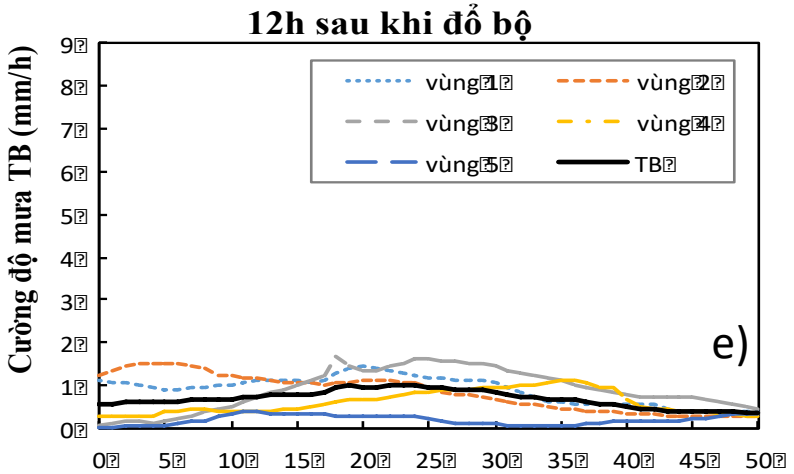
The WN-1 asymmetric rainfall component (M_1)can be represented by:

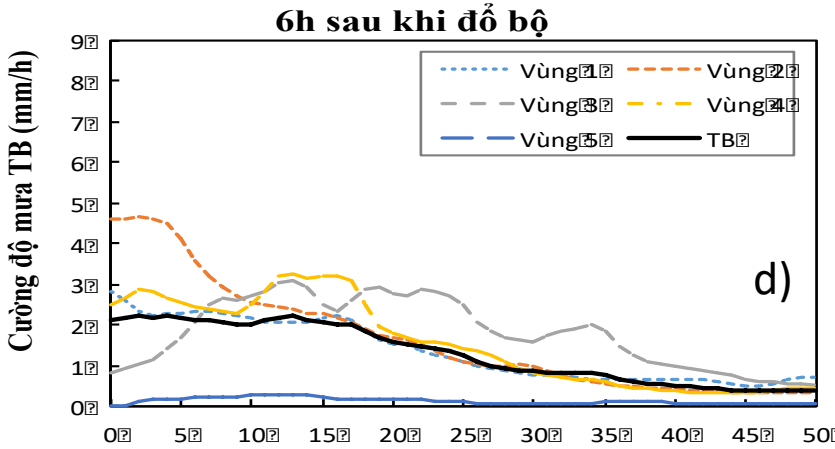
$$M_1 = [a_1 \cos(\theta) + b_1 \sin(\theta)]/R.$$

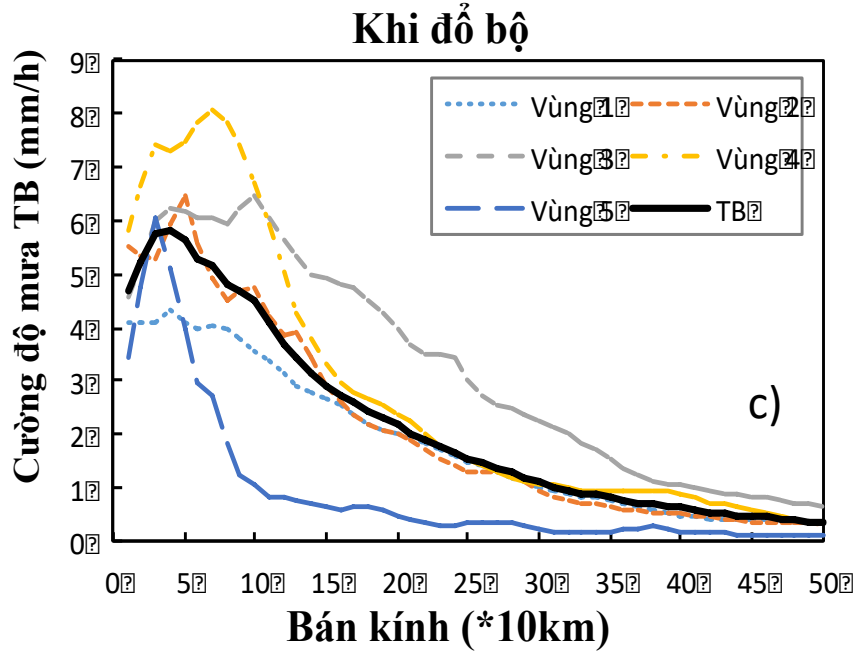
Characteristics of rainfall distribution induced by landed TCs divided by landing regions





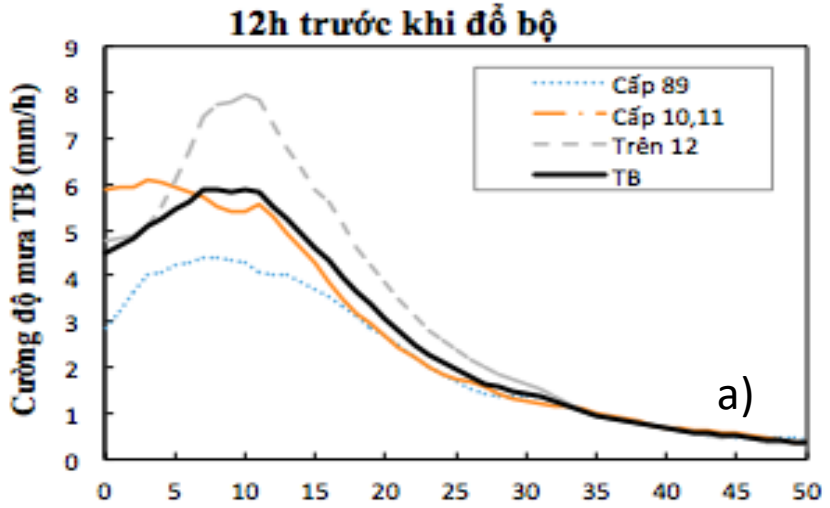


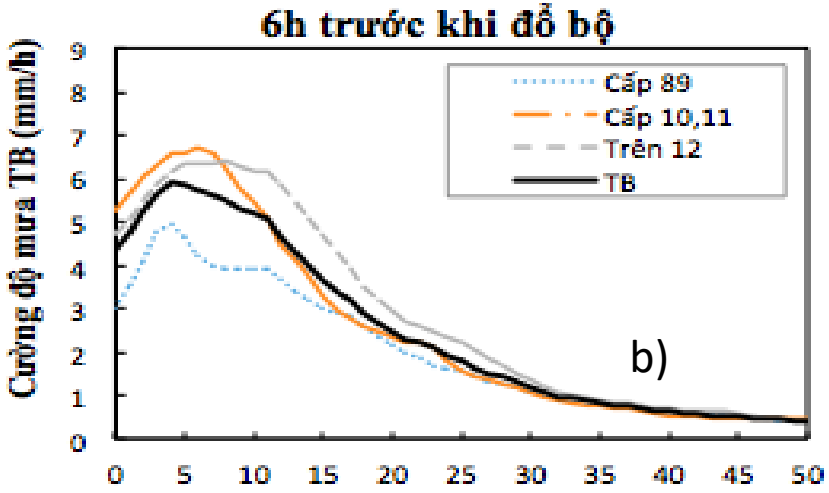


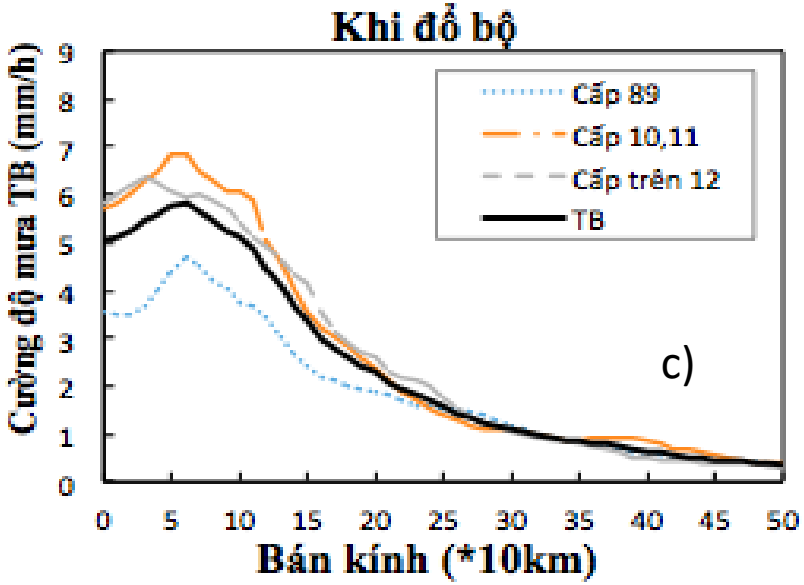


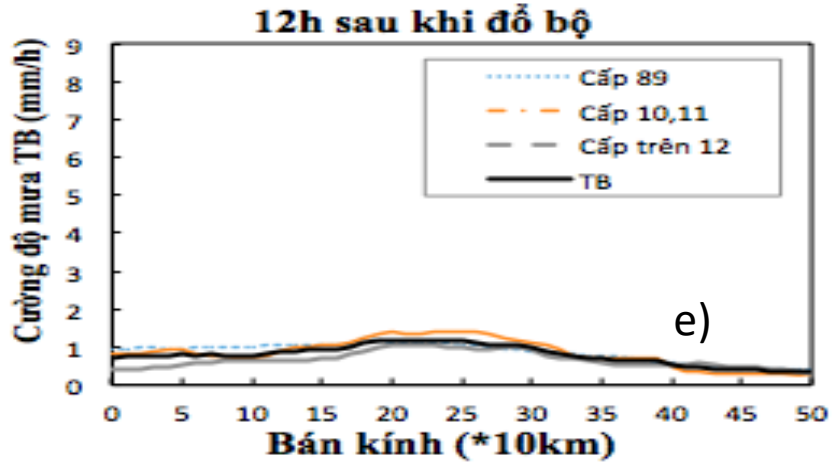
Radial profiles of the azimuthally averaged rain rates in TCs making landfall in different regions

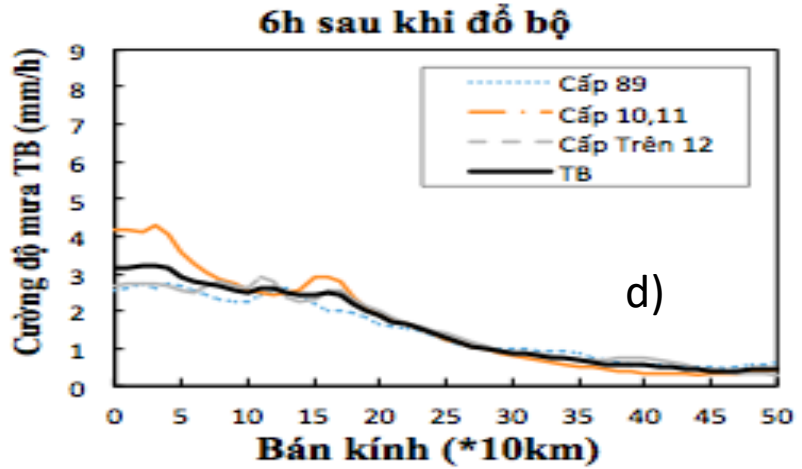
Characteristics of rainfall distribution induced by TCs making landfall divided by TC intensity







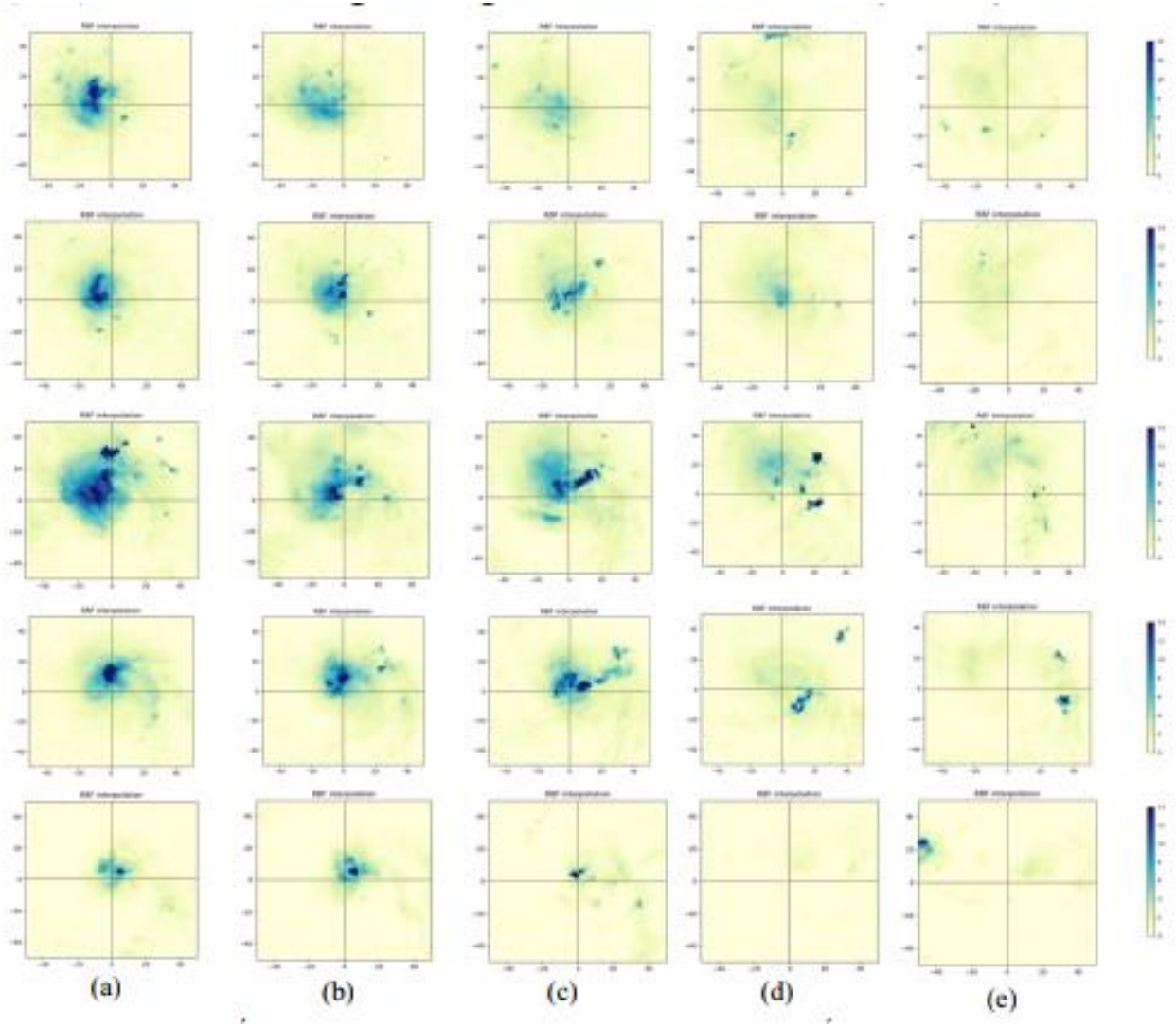




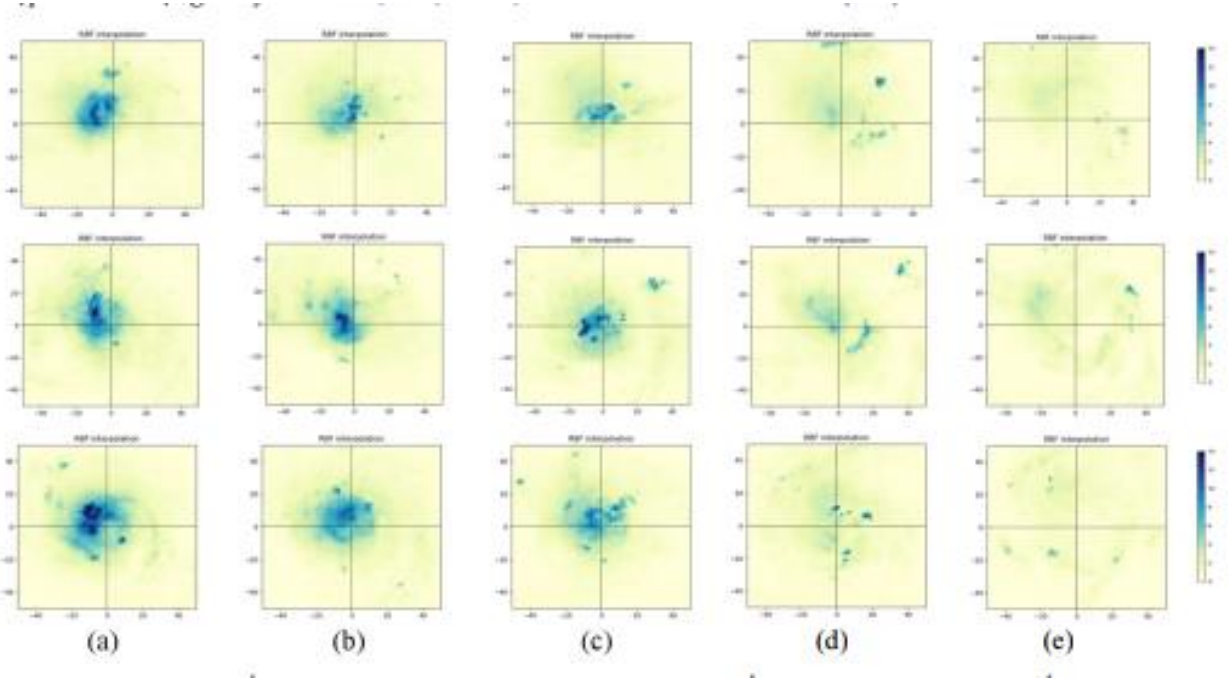
Radial profiles of the azimuthally averaged rain rates in TCs making landfall divided by TC categories

Distribution of rain intensity (mm/h) from the center of the storm to a radius of 500km

Divided by regions: from Region 1 to Region 5 (top to bottom)



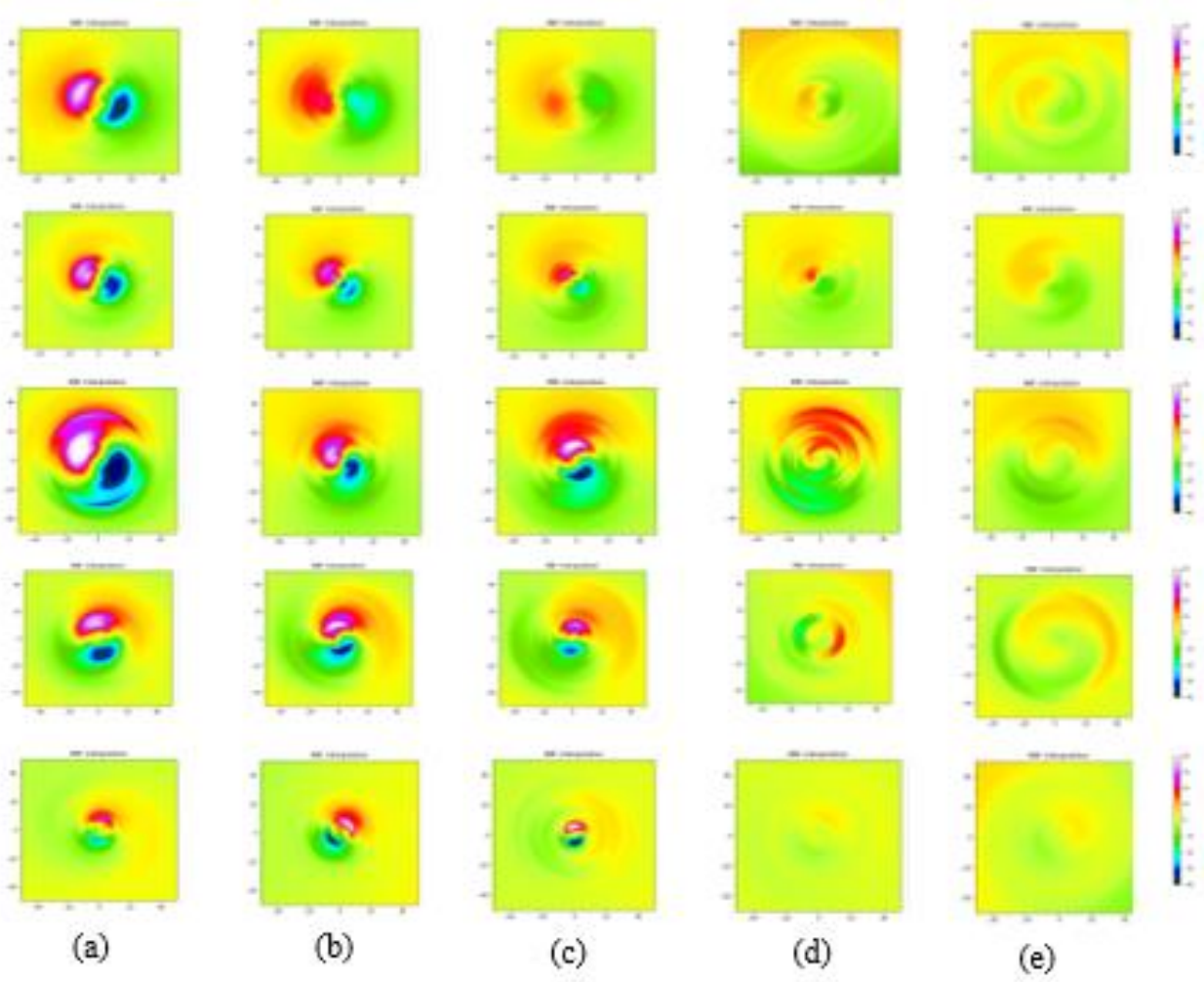
Divided by TC categories
TS (top), STS (middle), TYP (bottom)



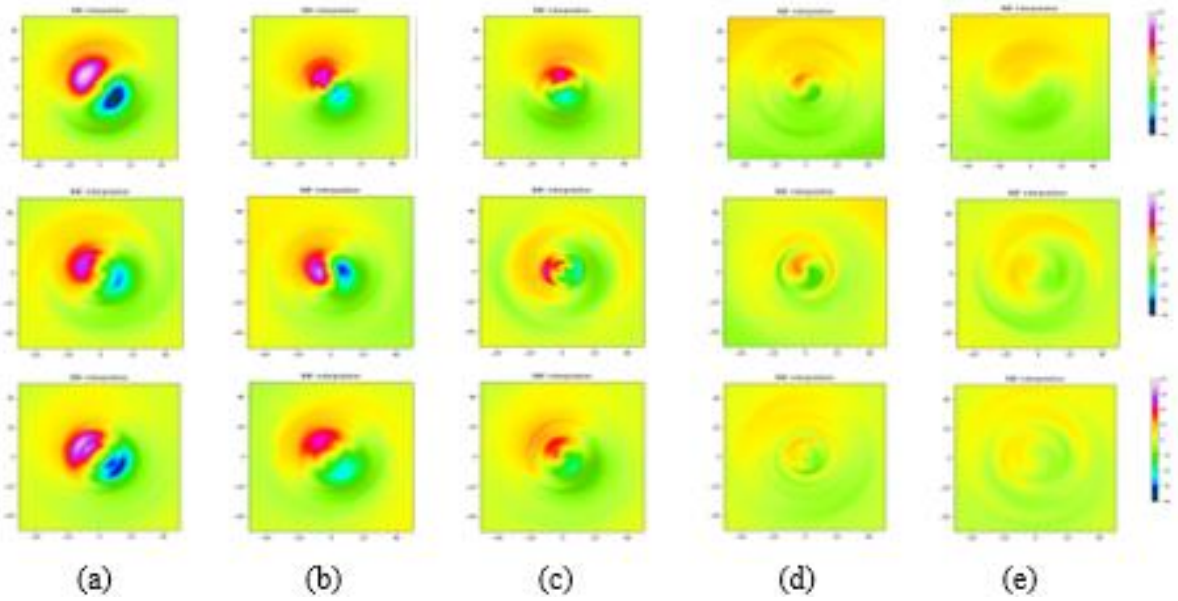
(a) 12h before landfall (b) 6h before landfall ; (c) at landfall;
(d) 6h after landfall; (e) 12h after landfall

The WN-1 rainfall asymmetry as a function of the distance from the TC center

Divided by regions: from Region 1 to Region 5 (top to bottom)



Divided by TC categories
TS (top), STS (middle), TYP (bottom)



(a) 12h before landfall (b) 6h before landfall ; (c) at landfall; d) 6h after landfall; (e) 12h after landfall

REMARKS

- The highest rainfall is associated with TC landing on the region III, with the peak of rain is approximately 100-150 km from the TC center
- Before landing, the distribution of maximum rainfall is located in the front left of TC moving direction when TCs make a landing on regions (I, II, III), while in the front right when TCs make a landing on regions (IV and V)
- While landing, the distribution of maximum rainfall is mostly turning clockwise, except for TC in the region (I) having anti-clockwise turning of maximum rainfall location
- There appeared an opposite turning of maximum rainfall location for different TC categories: clockwise for TS and TYP, while anti-clockwise for TST

3.3 SELECTIVE ENSEMBLE FORECAST OF TC-INDUCED RAINFALL USING GSMaP DATA

METHOD

Selective ensemble forecasting using GSMAP data is to improve ensemble forecast by selecting the best member when comparing with GSMaP by pattern matching method

1. Pattern matching method:

- ***Pattern Similarity by statistics***
  - Pixel-by-pixel verification methods: MAE or correlation
  - **Structural similarity index measure (**SSIM) AND **Cosine Similaritv** COSIN

$$\mathrm{SSIM(x,y)}=[l(x,y)]^{\alpha}\,.\,[c(x,y)]^{\beta}.\,[s(x,y)]^{\gamma}$$

$$\text{cosine similarity} = S_C(A,B) := \cos(\theta) = \frac{\mathbf{A}\cdot\mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}},$$

- ***K-means Clustering***
- ***HAC - Hierarchical Agglomerative Clustering***

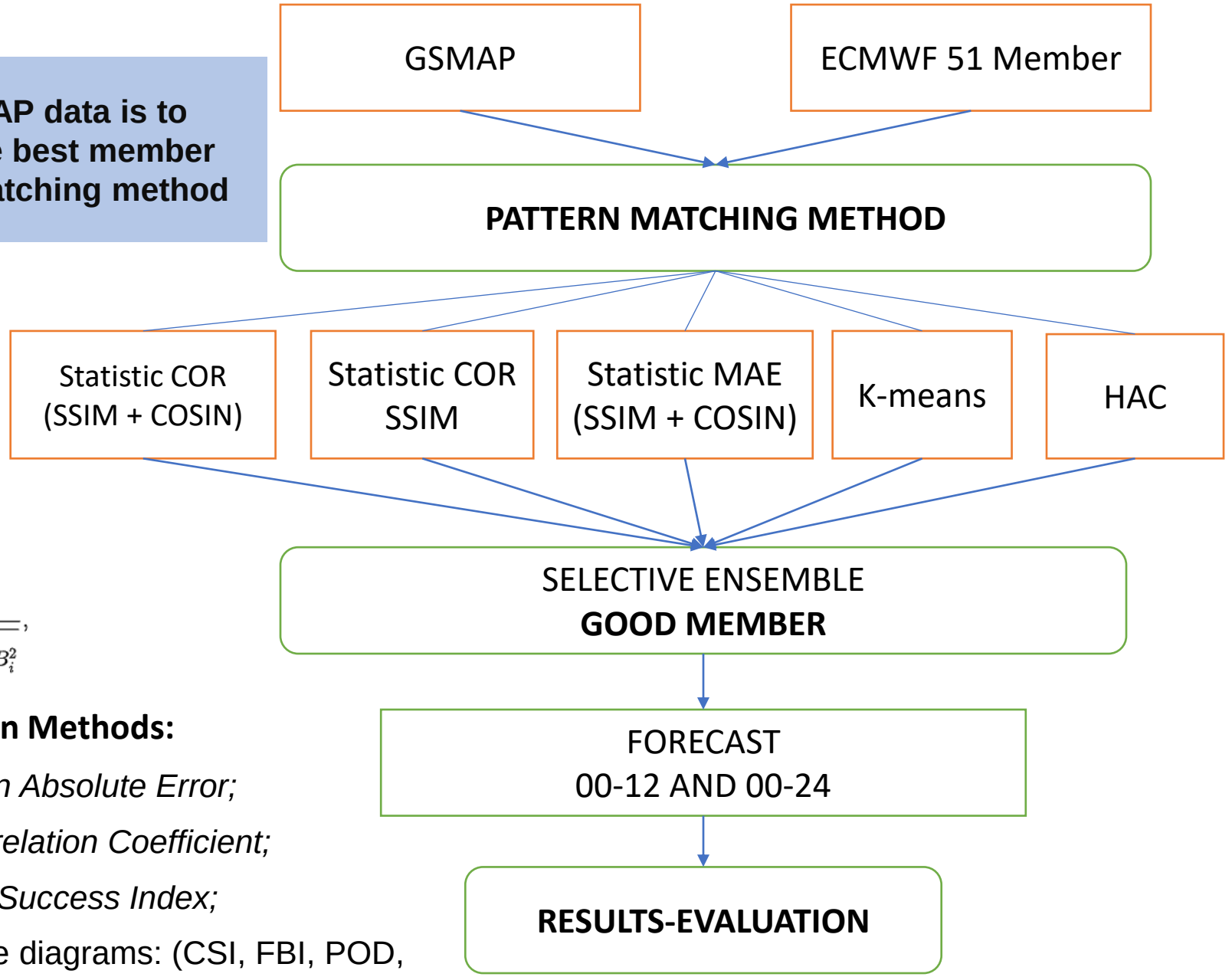
2. Evaluation Methods:

MAE – Mean Absolute Error;

CORR- Correlation Coefficient;

CSI-Critical Success Index;

Performance diagrams: (CSI, FBI, POD, FAR)



Selective ensemble forecasting using GSMAP data

Selection of SSIM and COSINSIM weights for the statistic method

Calculation of SSIM and Cosine similarity values 00-06

value = ssmi_w x ssmi + cosine_w x cosine

Repeat step: 0.1

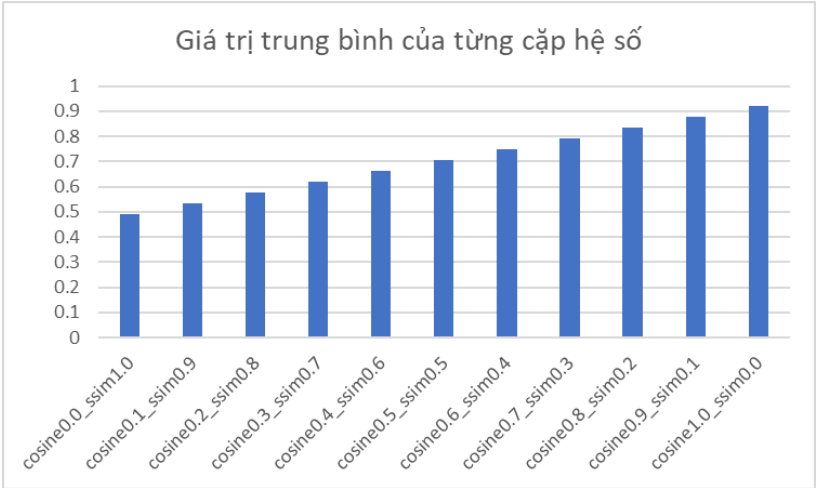
Calculate the value for each pair of weights

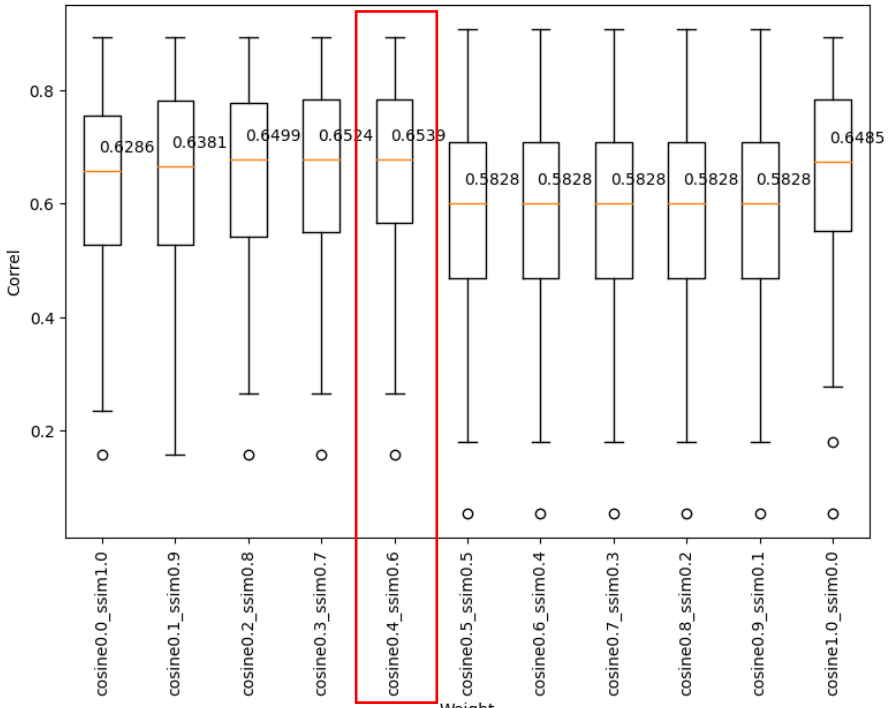
1. Iterate the weights and calculate the sum of SSMI and COSINE
2. Select members whose value is greater than the threshold

- ENS MEAN of SELECTED_MEMBERS
- CORREL with GSMAP

Using BOX PLOT to find the best of weight

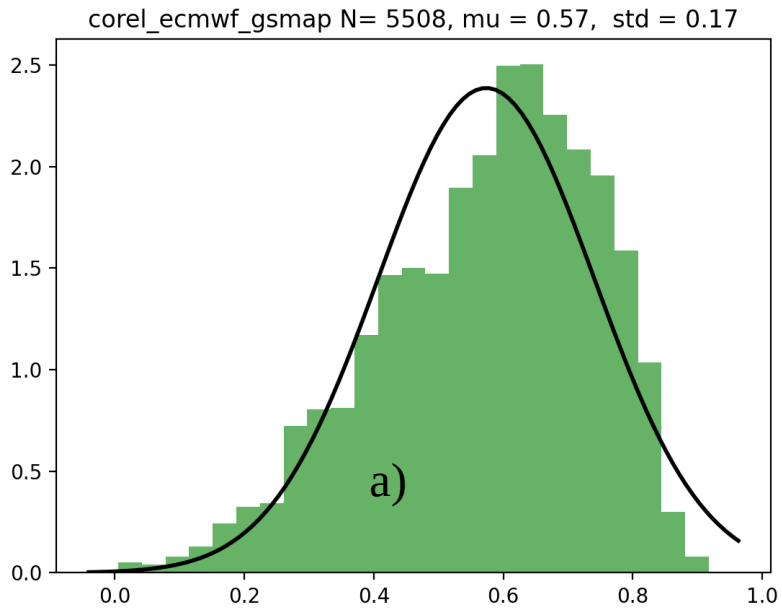
=> The weight: ssmi_w=0.6 , cosine_w=0.4

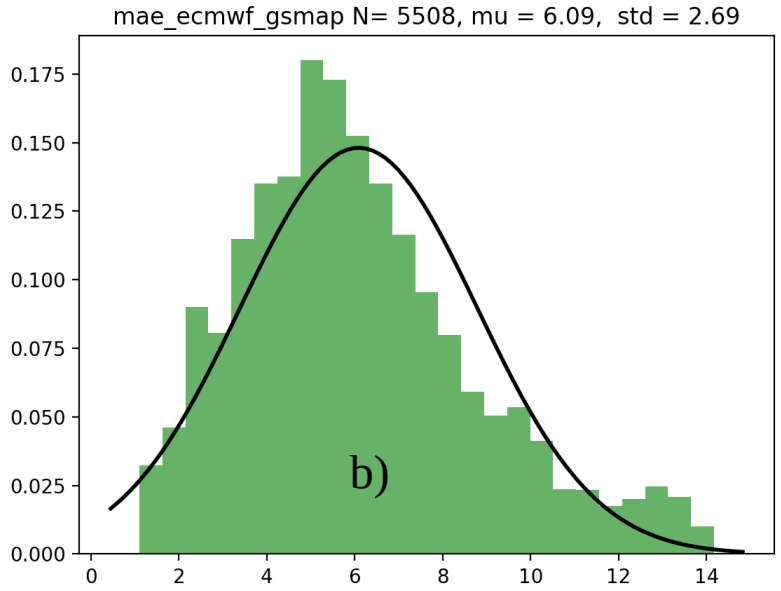


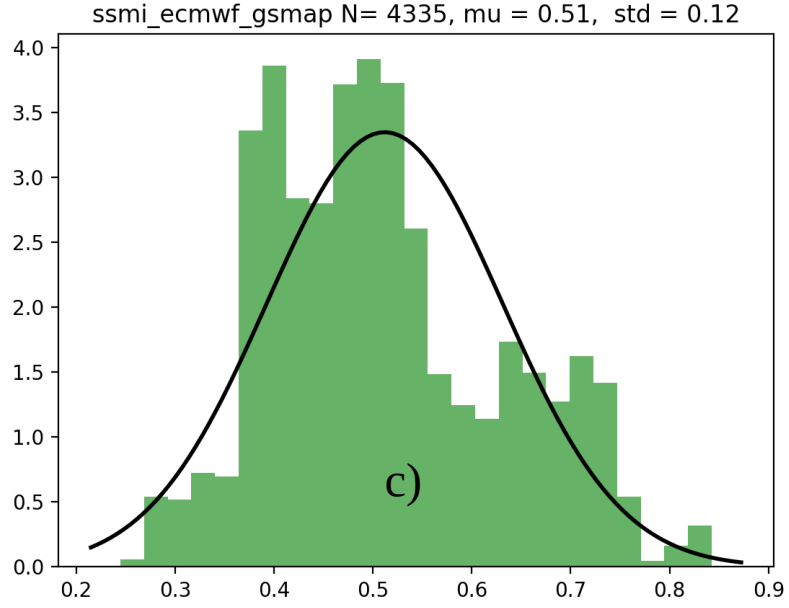


Selective ensemble forecasting using GSMAP data

Threshold selection





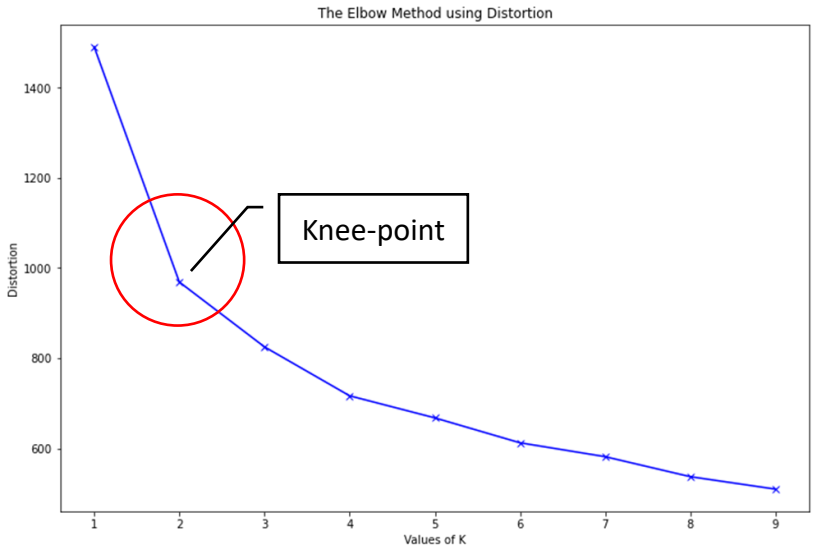


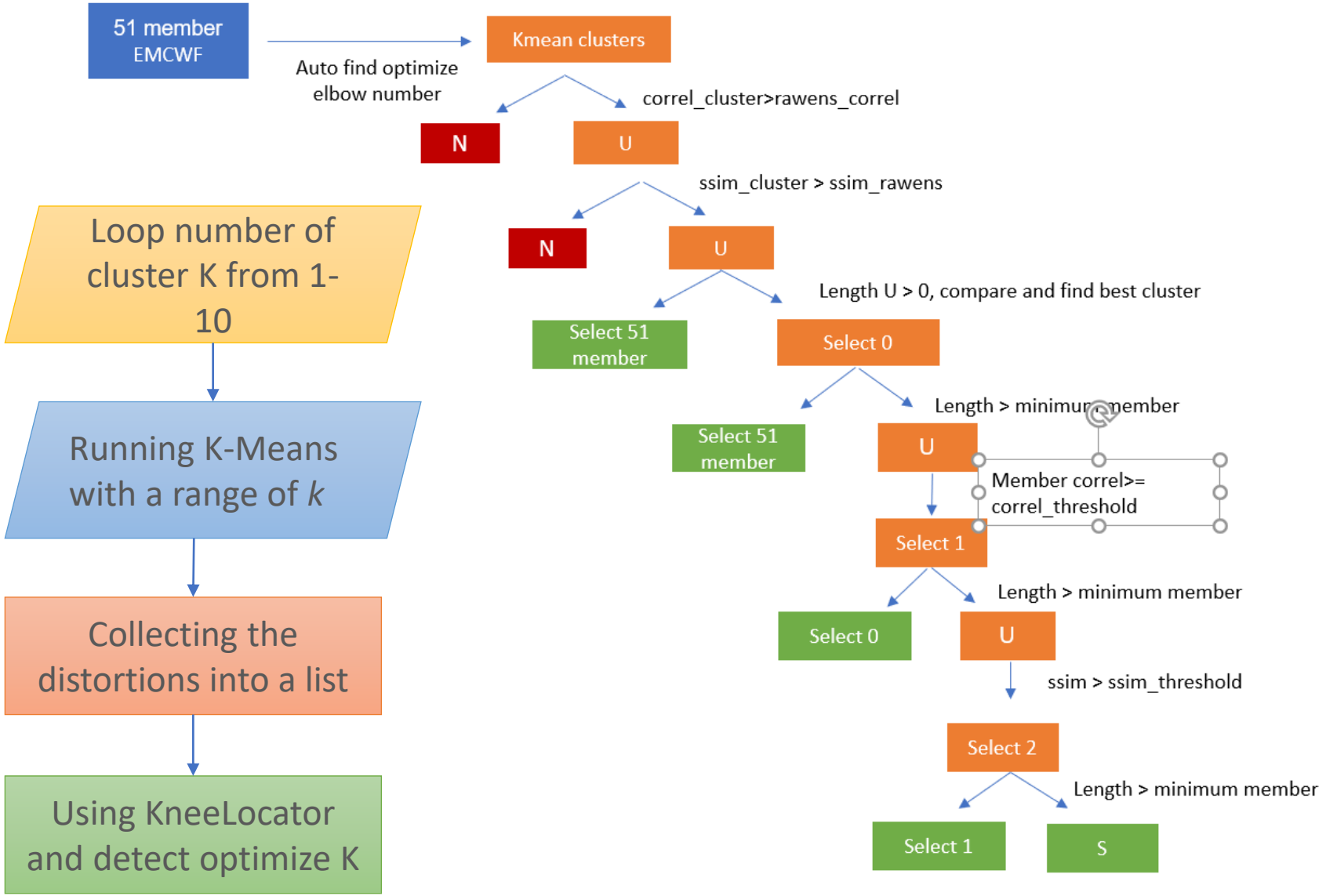
Graph of the probability density function, representing correlation value between GSMAP and ECMWF at time 00-06 of a) Correlation, b) MAE, c) SSIM index. Sample capacity is the number of cases X 51 = 5508

Selective ensemble forecasting using GSMAP data

Kmeans – tree method

Finding the optimal number of clusters is an important part of this algorithm.
A commonly used method to find the optimal K value is **the Elbow Method.**





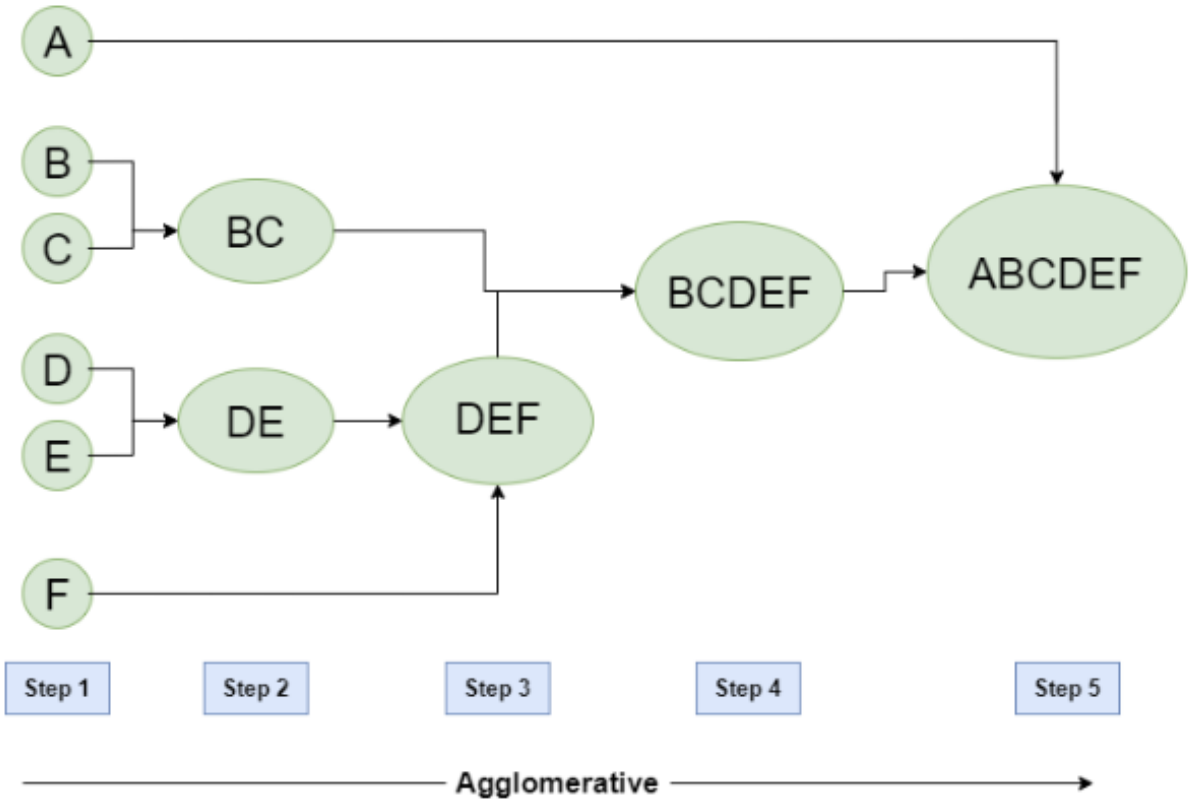
https://pypi.org/project/kneed/

Forecasting using GSMAP data and selective ensemble technique

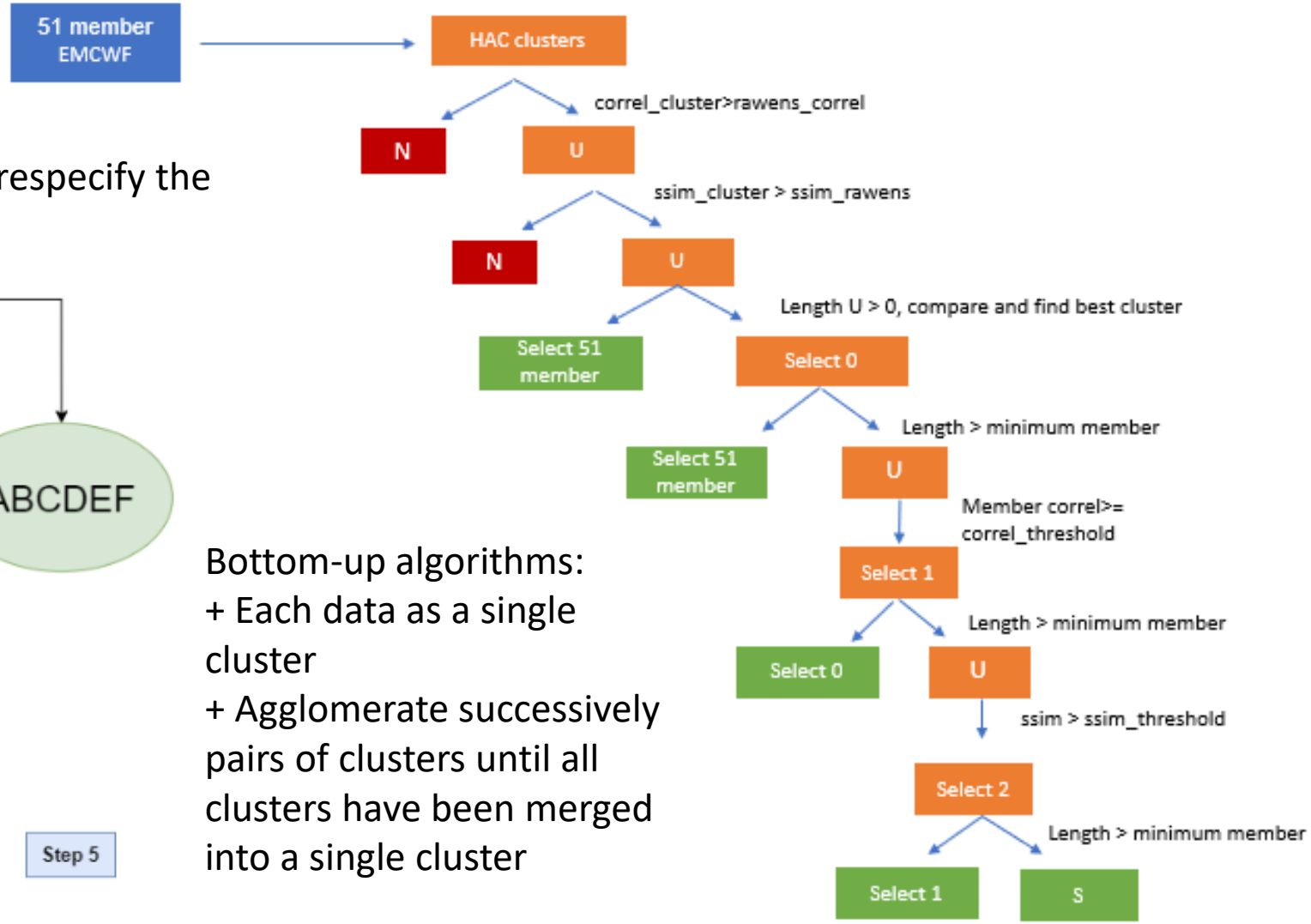
HAC – tree Method

HAC is known as the bottom-up approach

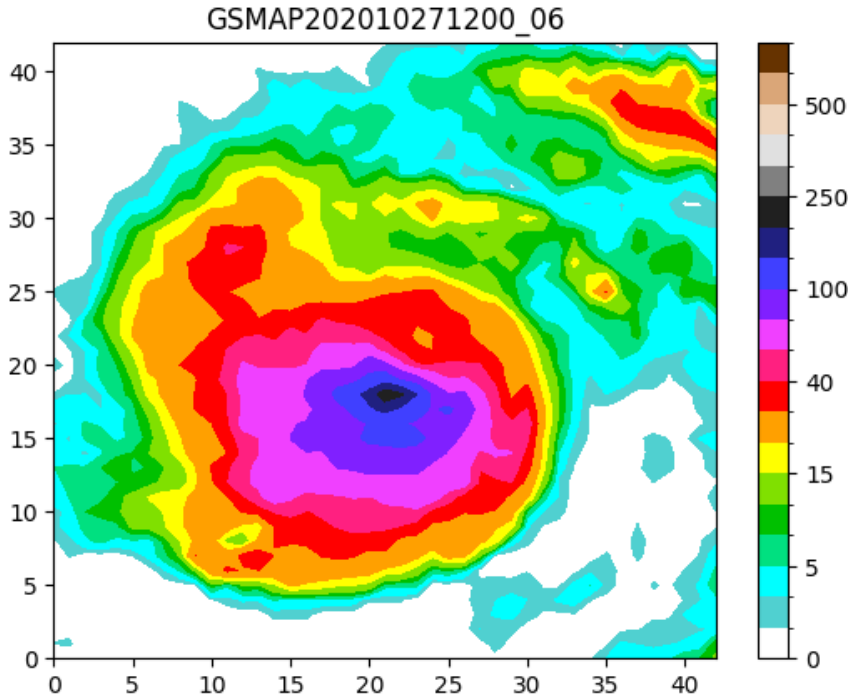
This clustering algorithm does not require us to prespecify the number of clusters.

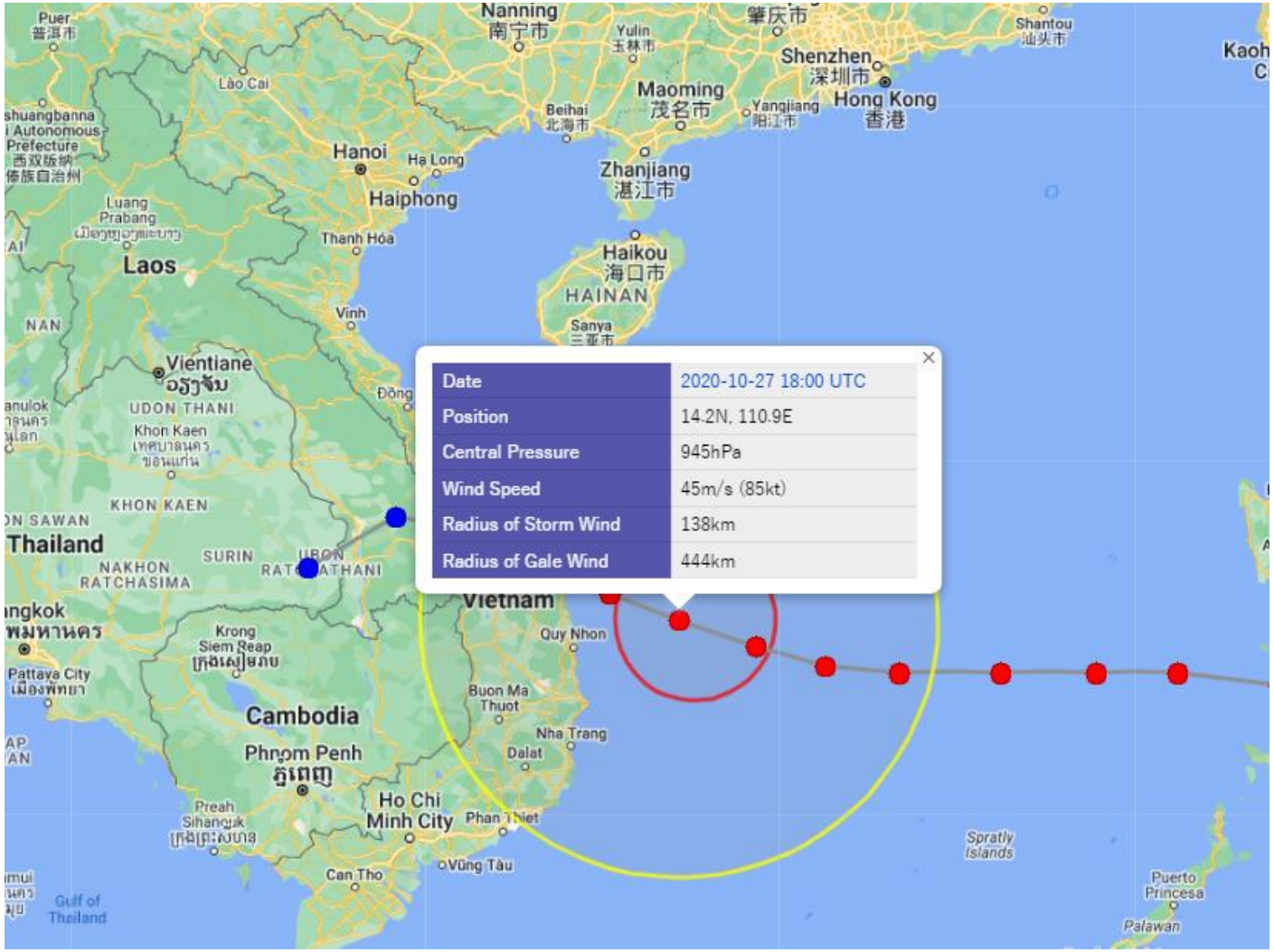


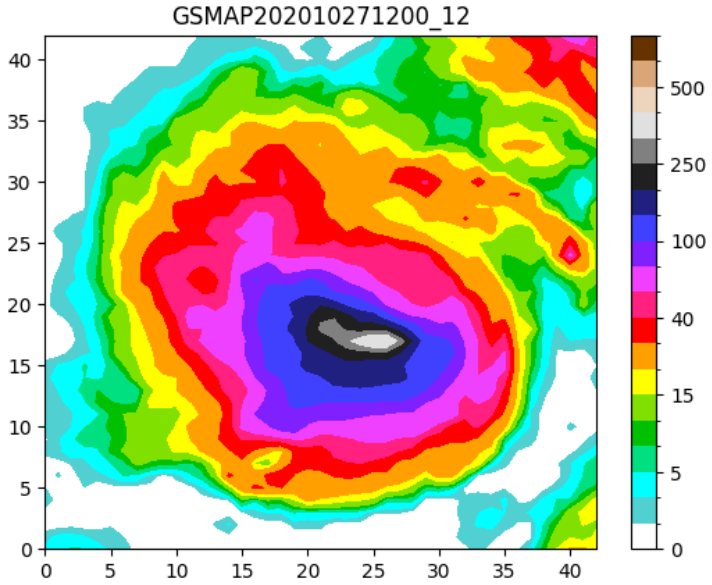
HAC – Clustering

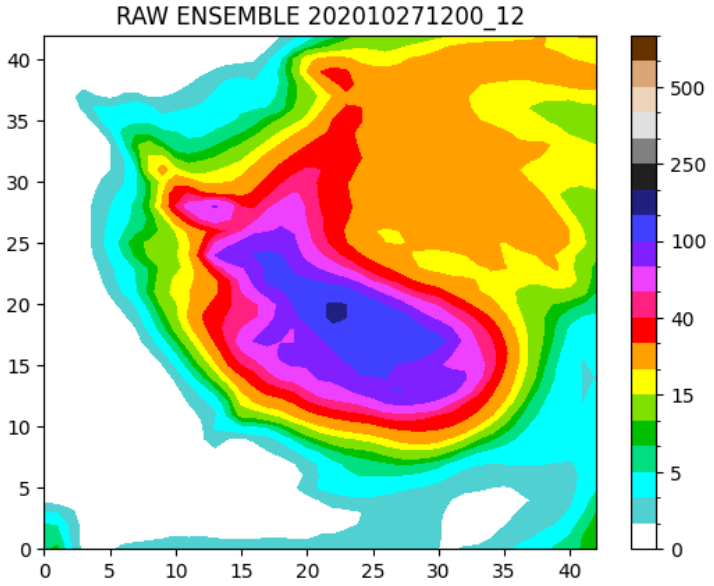


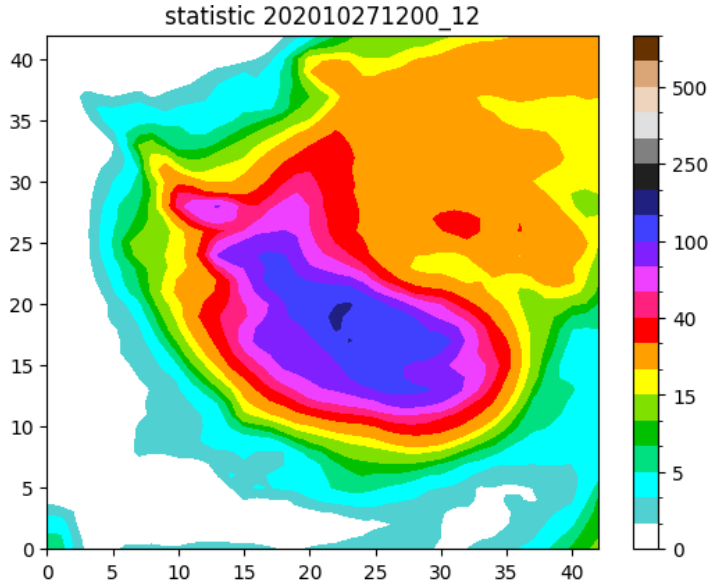
Bottom-up algorithms:
+ Each data as a single cluster
+ Agglomerate successively pairs of clusters until all clusters have been merged into a single cluster

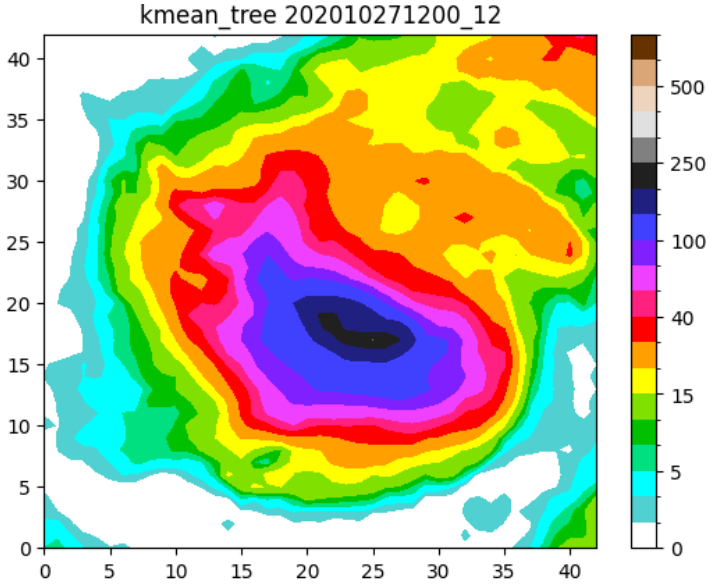

GSMAP202010271200_06
0
5
10
15
20
25
30
35
40
500
250
100
40
15
5
0

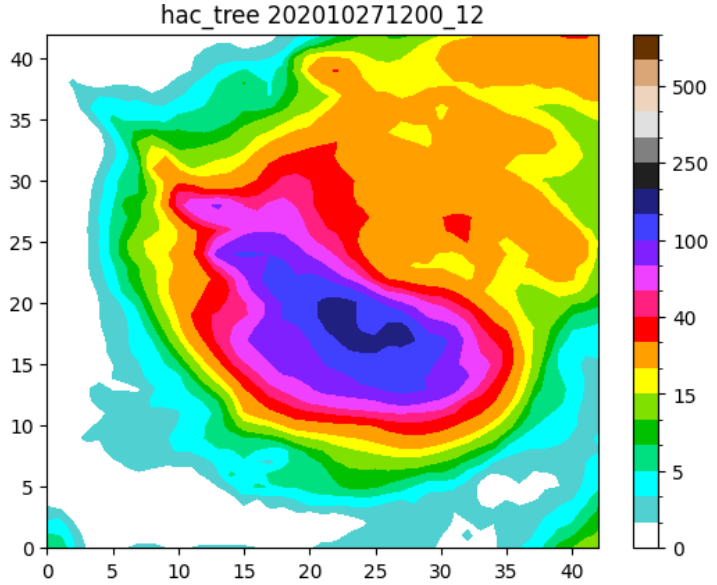

Nanning
南宁市
Yulin
玉林市
Shenzhen
深圳市
Hong Kong
香港
Maoming
茂名市
Zhanjiang
湛江市
Haikou
海口市
HAINAN
Sanya
Hanoi
Haiphong
Hạ Long
Laos
Thanh Hóa
Vinh
Vientiane
UDON THANI
Khon Kaen
KHON KAEN
Thailand
NAKHON RATCHASIMA
SURIN
Vietnam
Quy Nhon
Cambodia
Phnom Penh
Buon Ma Thuot
Nha Trang
Dalat
Ho Chi Minh City
Vũng Tàu
Can Tho
Gulf of Thailand
Spratly Islands
Puerto Princesa
Palawan
Pattaya City
Krong Siem Reap
Lào Cai
Beihai
Yangjiang
Date 2020-10-27 18:00 UTC
Position 14.2N, 110.9E
Central Pressure 945hPa
Wind Speed 45m/s (85kt)
Radius of Storm Wind 138km
Radius of Gale Wind 444km

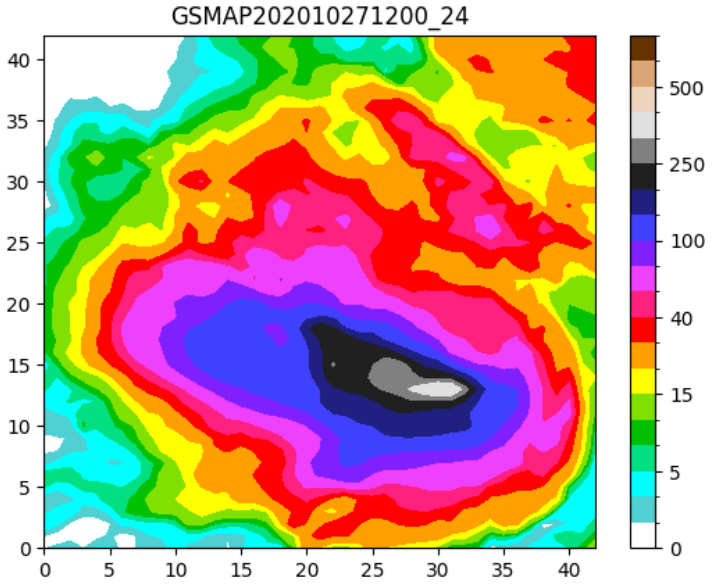

GSMAP202010271200_12
0 5 10 15 20 25 30 35 40
500 250 100 40 15 5 0

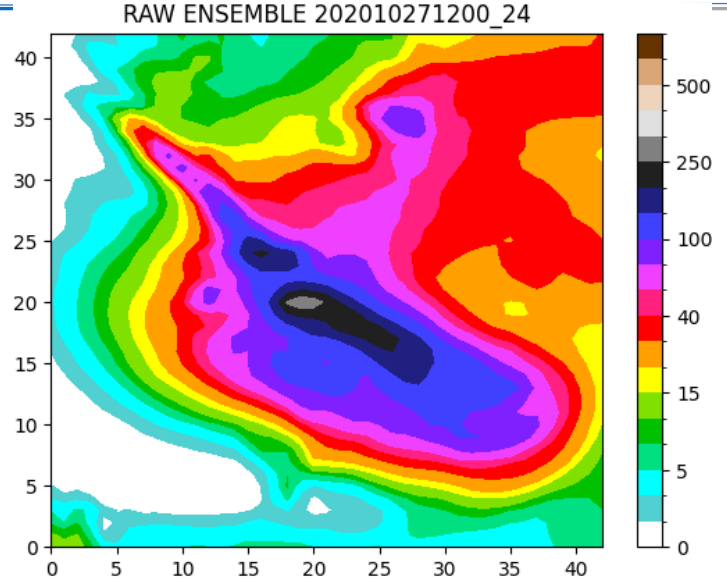

RAW ENSEMBLE 202010271200_12
0 5 10 15 20 25 30 35 40
500 250 100 40 15 5 0

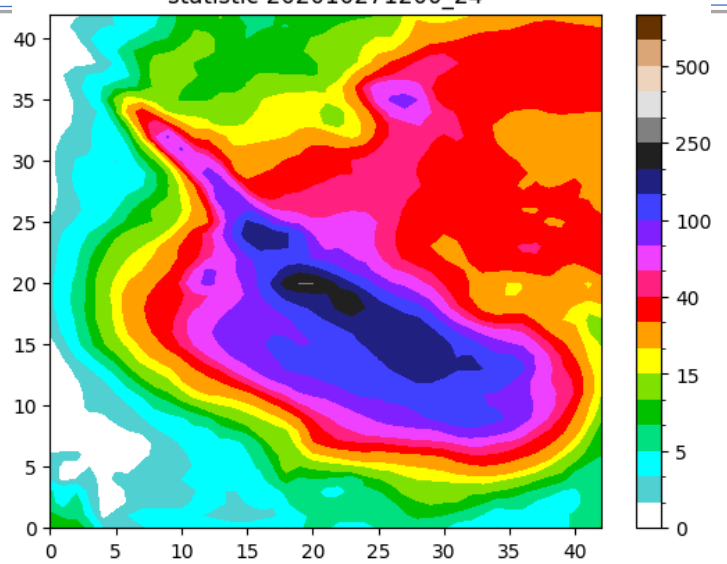

statistic 202010271200_12
0 5 10 15 20 25 30 35 40
500 250 100 40 15 5 0

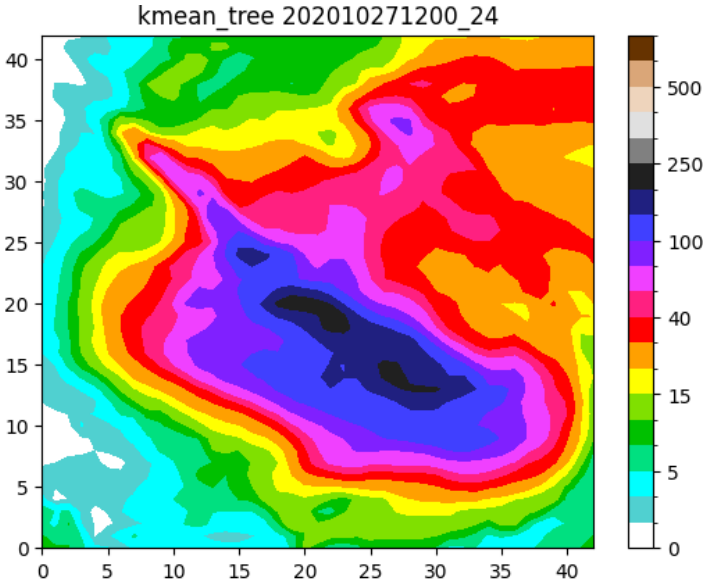

kmean_tree 202010271200_12
0 5 10 15 20 25 30 35 40
500 250 100 40 15 5 0

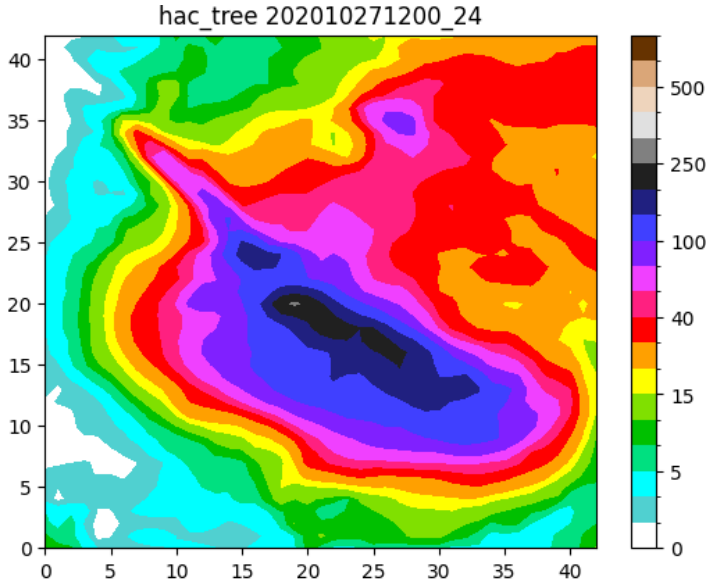

hac_tree 202010271200_12
0 5 10 15 20 25 30 35 40
500 250 100 40 15 5 0


GSMAP202010271200_24


RAW ENSEMBLE 202010271200_24

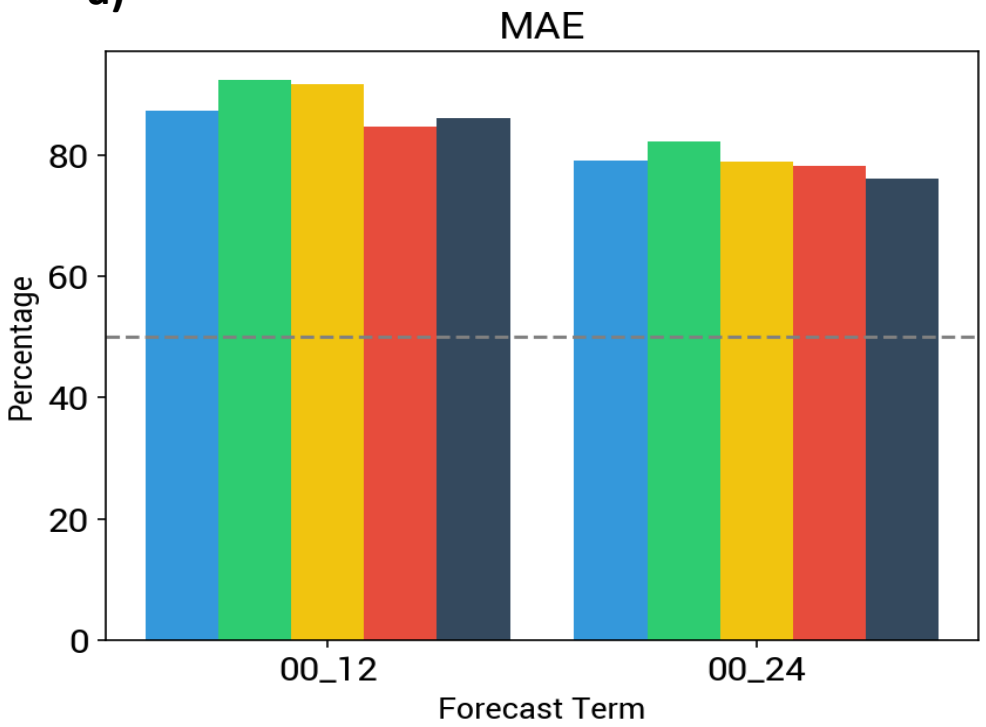

statistic 202010271200_24


kmean_tree 202010271200_24

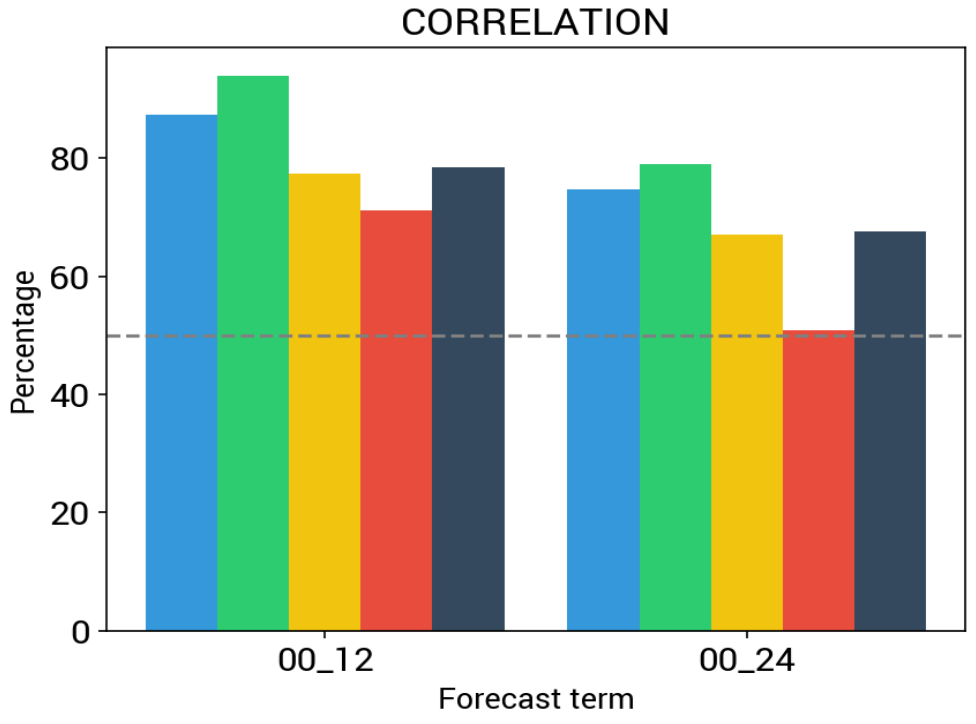

hac_tree 202010271200_24

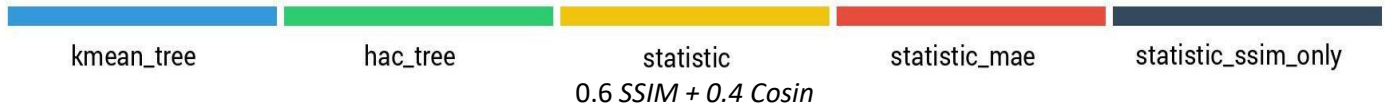
DETERMINISTIC FORECASTS RESULTS (1)

a)



b)





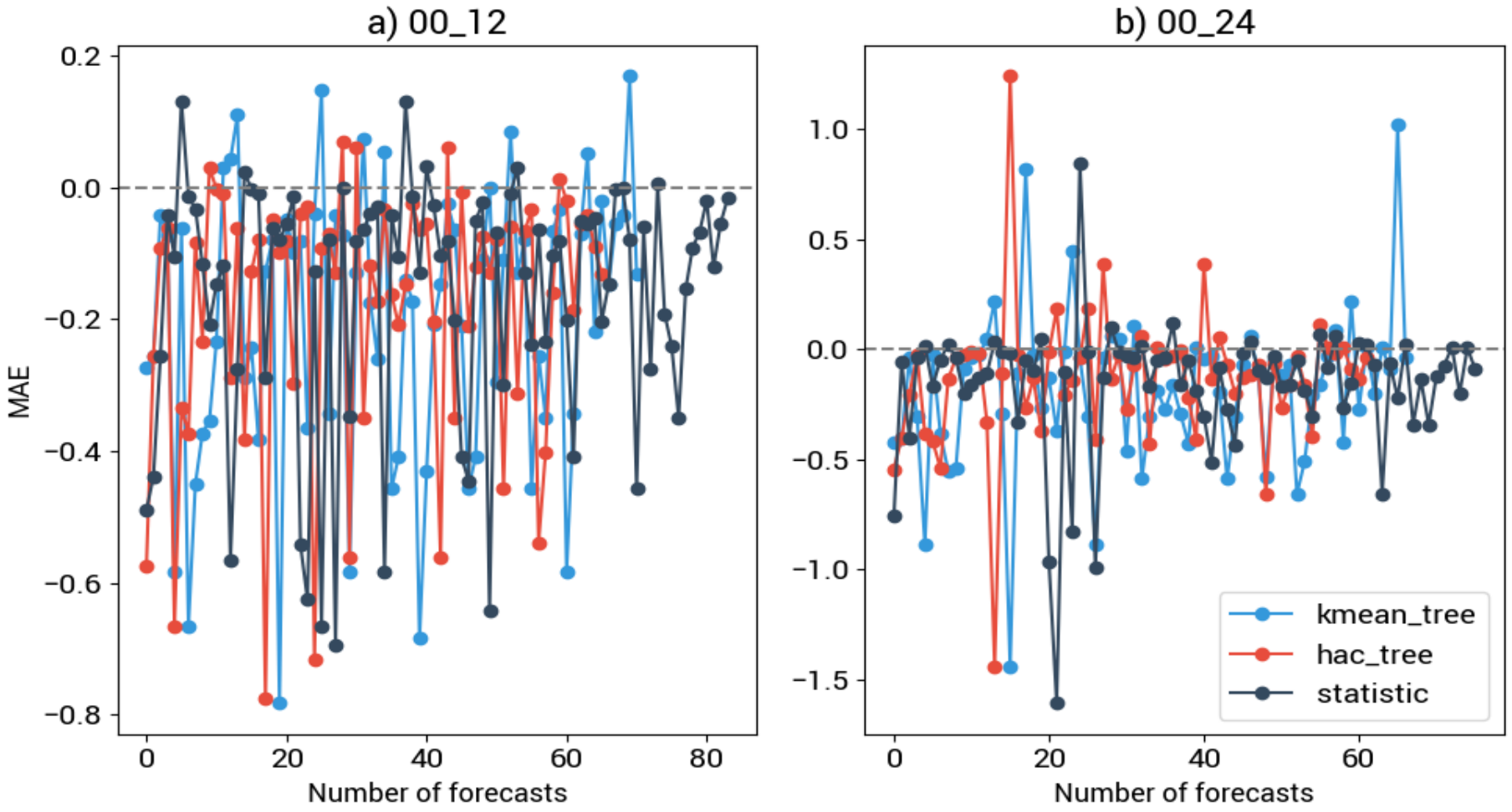
Percentage of times the forecast is better than rawens at 00-12 and 00-24: a) MAE b) CORRELATION

- The results show that the options give better results in both indicators and at 00-12 better than 00-24
- The HAC option is the one that gives the best results
- With the statistical methods show that CORRELATION using the combination SSIM+ COSIN is the best

FORECASTS EVALUATION

- *MAE at forecast time 00-12 improves more than the term 00-24*
- *Most of the time the forecast has negative results, indicating an improvement over raw ens.*

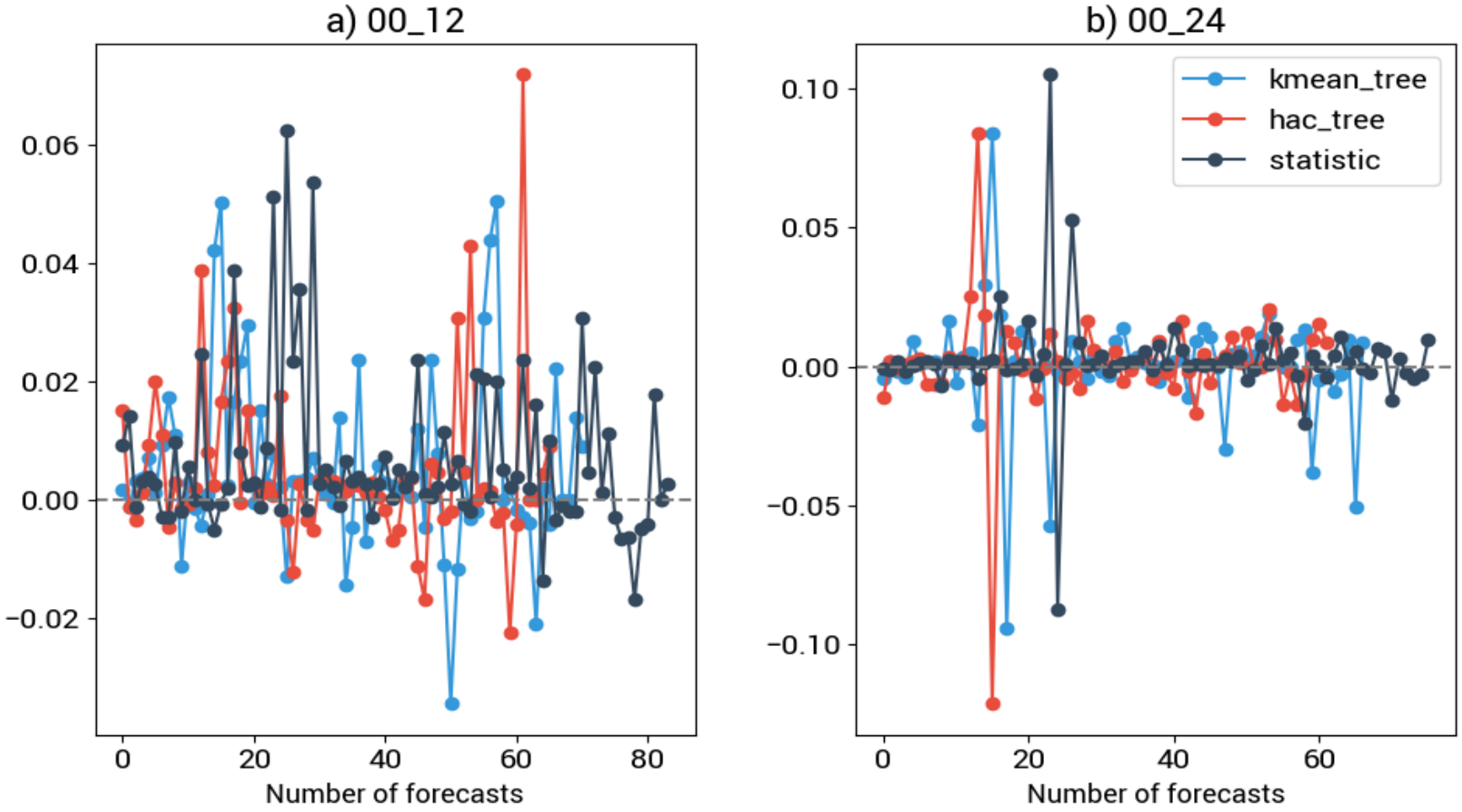
Statistical results on quantity:
MAE (Pattern matching) – MAE (raw ens)



DETERMINISTIC FORECASTS RESULTS (4)

Statistical results on quantity
Calculate:
Correlation (Pattern matching) –
Correlation (raw ens)

Correlation

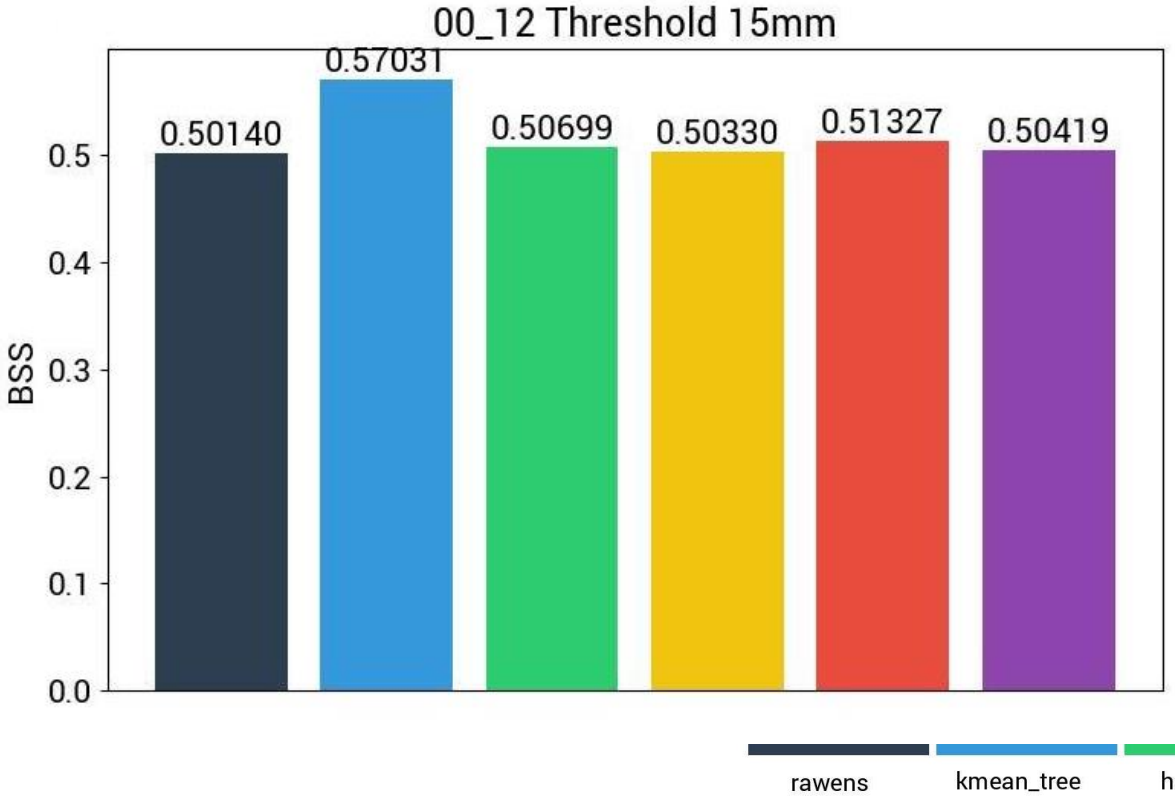


Forecasting using GSMAP data and selective ensemble technique

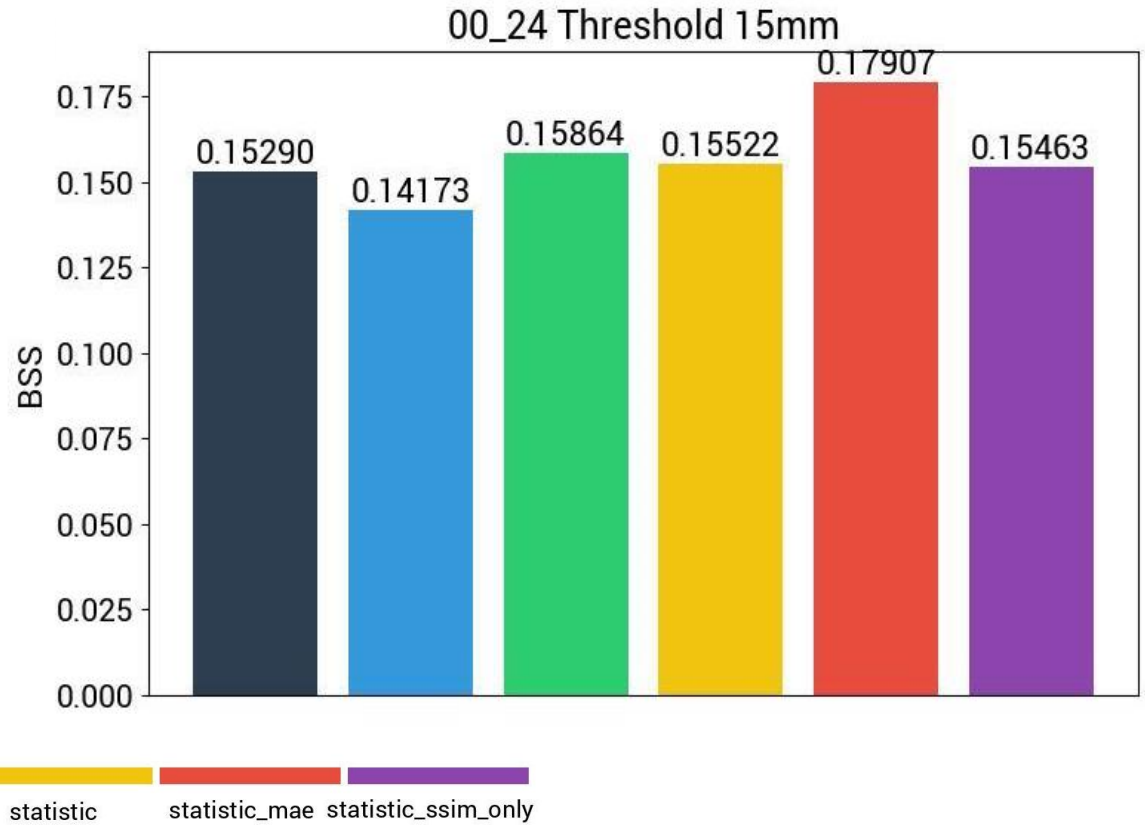
PROBABILITY FORECAST RESULTS (1)

Brier Skill Score

BSS index in Leadtime 00-12



BSS index in Leadtime00-24



The K-mean option gives the best results,
the other methods have a slight improvement over raw ens

the statistic_mae method gives the best results, and
K-mean gives the worst error, lower than raw_ens.

SUMMARY

1) Evaluation of TRMM/GSMaP provides a similar bias trend for TC-induced rains

- There are the significant difference about the distribution and magnitude of bias between affected regions.
- TRMM and GSMAP are over estimasted in the regions within 500km from TC core. It except for the radius of 200km in the right side of TC event effecting to the AR3.

2) Asymmetric rainfall analysis shows that the average rainfall intensity as well as the distribution of maximum rainfall areas are significantly different in landing regions and depending on TC intensity.

- The distribution of the maximum rainfall is concentrated in the front left of the storm's movement in regions (1,2,3), except for the storms in regions 4 and 5, where the maximum rainfall is skewed to the right of moving direction.
- Mostly clockwise turning of maximum rainfall distribution for TC landing, except for region 1

3) Selective ensemble forecast using GSMaP data shows improvement in percentage of forecast times over raw-ensemble in term of evaluation criterie

- The K-mean option gives the best results, the other methods have a slight improvement over raw-ensemble

Trân trọng cảm ơn!